AUGUST
1955
VOLUME 19
NUMBER 4

JOURNAL OF CONSULTING PSYCHOLOGY

AMERICAN PSYCHOLOGICAL ASSOCIATION

JOURNAL OF CONSULTING PSYCHOLOGY

Published by the American Psychological Association, Inc.
at Prince and Lemon Sts., Lancaster, Pa.
and 1333 Sixteenth St. N.W., Washington, D.C.

Edited by Laurance F. Shaffer

Teachers College, Columbia University

Advisory Editors: Ann Magaret Garner, *University of Illinois*. William A. Hunt, *Northwestern University*. E. Lowell Kelly, *University of Michigan*. Morris Krugman, *Board of Education, New York, N. Y.* Bertha M. Luckey, *Board of Education, Cleveland, Ohio*. Fred McKinney, *University of Missouri*. Anne Roe, *New York, N. Y.* Carl R. Rogers, *University of Chicago*. Nevitt Sanford, *University of California*. Edward Joseph Shoben, Jr., *Teachers College, Columbia University*. J. R. Wittenborn, *Rutgers University*.

Lorraine Bouthilet, *Managing Editor*

The *Journal of Consulting Psychology* is the clinical journal of the American Psychological Association. It is devoted primarily to the publication of original research relevant to psychological diagnosis, psychotherapy and counseling, personality, and the dynamics of behavior. Case studies, theoretical contributions, descriptions of clinical techniques, and discussions of the professional problems of clinical psychology also appear in its pages.

Editorial Office

Manuscripts submitted for publication and correspondence on editorial matters should be sent to the Editor, Laurance F. Shaffer, 525 West 120th St., New York 27, N. Y. All manuscripts should be submitted *in duplicate* to facilitate referral to the Advisory Editors. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

Style of manuscripts. Manuscripts should be prepared in the style described in the Publication Manual of the American Psychological Association (*Psychological Bulletin*, 1952, 49, 389-449). The Manual may be purchased from the Washington office of the Association for $1.00. Otherwise acceptable manuscripts which depart from the standards of the Manual are returned for revision and retyping.

Brief Reports. A one-page condensation of an acceptable research report may be given early publication without charge if the author agrees (*a*) to make 100 mimeographed copies of the full report available to those who request it, (*b*) to send a copy of the full report to the Editor for deposit in the American Documentation Institute, and (*c*) not to have the full report published elsewhere. For detailed instructions on how to prepare a Brief Report, see the back of the *Contents* page of this Journal or write to the Editor.

Early Publication. Authors may secure the early publication of an acceptable article in the first issue to go to press after its acceptance by arranging for the article to be printed as extra pages, increasing the number of pages that subscribers receive. The charge to the author for such early publication is $20.00 per printed page.

Reprints. Authors receive fifty reprints gratis of all articles except early publication articles and Brief Reports. Additional quantities may be ordered when returning proof.

Business Office

Communications concerning subscriptions, change of address, claims for the nonreceipt of a number, advertising, and other business matters, should be sent to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D.C. Address changes must reach the subscription office by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

The *Journal of Consulting Psychology* is published bimonthly in February, April, June, August, October, and December, the yearly volume comprising approximately 480 pages. The subscription per year is $7.00, foreign $7.50, single number $1.50.

Entered as second-class matter, March 21, 1955, at the post office at Lancaster, Pa., under the act of March 3, 1879. Acceptance for mailing at special rate of postage provided for in Section 538, Act of February 25, 1925, authorized May 27, 1940.

Copyright, 1955, by the American Psychological Association, Inc.

JOURNAL OF CONSULTING PSYCHOLOGY

August, 1955 Vol. 19, No. 4

Contents

Changes in Psychoneurotic Patients With and Without Psychotherapy: *Frank Barron and Timothy F. Leary* - 239

Brief Psychotherapy and Enuresis: *Daniel H. Harris, Richard W. Firestone, and Carl M. Wagner* - 246

The Measurement of Depth of Interpretation: *Norman I. Harway, Allen T. Dittmann, Harold L. Raush, Edward S. Bordin, and David Rigler* - - - - - - - - - - - - 247

Responses to Imagined Auditory Stimuli as Compared to Recorded Sounds: *David R. Stone* - 254

Clinical Statements as Scientific Propositions and Social Decisions: *Frank M. du Mas* - - 255

Delayed Gratification and Ego Development: Implications for Clinical and Experimental Research: *Jerome L. Singer* - 259

Testing Behavior Hypotheses with the Rorschach: An Exploration in Validation: *Arthur E. Eschenbach and Edgar F. Borgatta* - 267

Predicting Success on the Member-Employee Rehabilitation Program: *Bernard A. Stotsky* 274

Relations Among Several Objective Measures of Anxiety Under Different Conditions of Motivation: *Anthony Davids* - 275

A Further Investigation of the Relationship Between Manifest Anxiety and Intelligence: *Allen D. Calvin, Paul B. Koons, Jr., Joseph L. Bingham, and Howard H. Fink* - - 280

The Correlation of Intelligence and Neurological Findings on Twenty-Two Patients Autopsied for Brain Damage: *Robert S. Morrow and Joseph C. Mark* - - - - - - - 283

The Awareness of Hostile Feelings in Patients by Their Nurses: *Sol Charen* - - - - - - - 290

Psychological Changes over a Three-Year Period Following Bilateral Prefrontal Lobotomy: *Isidor W. Scherer, John F. Winne, and Robert W. Baker* - - - - - - - - - - - - 291

Diagnosis of Organicity by Means of Spiral Aftereffect: *A. Cooper Price and H. L. Deabler* 299

The Efficacy of Diagnostic Pattern Analysis with the Wechsler-Bellevue: *Jacob Cohen* - - 303

Specific Ability in IQ Change: *Charles T. Baker, Lester W. Sontag, and Virginia L. Nelson* 307

Rigidity on the Rorschach and Response to Intermittent Photic Stimulation: *Laverne C. Johnson and John A. Stern* - 311

New Books and Tests - 318

Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specifications given below.

2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.

3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.

4. Agrees not to submit the full report to another journal of general circulation.

Specifications

Brief Report. The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the author's lines, must not exceed 70 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style (1). Headings, tables, and references are avoided or, if essential, must be counted in the 70 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 70-line quota: [1]

[1] An extended report of this study may be obtained without charge from John Doe, 300 Market St., Prospect 6, Mass. (giving the author's full name and address), or for a fee from the American Documentation Institute. Order Document No. —— from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $—— for microfilm or $—— for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Extended report. The full report is prepared in the style specified by the *Publication Manual* (1), except that it may be typed with single spacing for economy in photoduplication by the ADI.

Reference

1. American Psychological Association. Council of Editors. Publication manual of the American Psychological Association. *Psychol. Bull.*, 1952, 49, 389–449.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Changes in Psychoneurotic Patients With and Without Psychotherapy[1]

Frank Barron
Institute of Personality Assessment and Research, University of California, Berkeley

and Timothy F. Leary
Kaiser Foundation Hospital, Oakland, California

Two tendencies seem constantly at work in living systems: the tendency to integrate a durable and stable arrangement of internal parts and external relationships, and the interacting opposite of this, the tendency to respond variably either to pressures arising from internal structure or to new external conditions in such a way as to break down old forms of organization and replace them with new ones.

In psychoneurosis, which is in part at least an attempt to stabilize matters internally and externally, one might expect also that forces arising internally and tending towards change would be at work. As Eysenck (3) has pointed out in his review of reports on the incidence of improvement under varying therapeutic regimes, the rate of spontaneous remission in untreated psychoneuroses is almost identical with the recovery rate reported when intensive psychotherapy has been applied. While Eysenck's analysis of these reported data has been challenged by Rosenzweig (9), and perhaps cannot be considered to be compelling in the light of Rosenzweig's criticisms, it is undoubtedly true that a substantial percentage of untreated psychoneuroses and psychoses make some sort of recovery.

[1] The present investigation is one of a series of studies sponsored by the Kaiser Foundation, Oakland, California, and supported in part by Research Grant MH–331 from the National Institute of Mental Health, Public Health Service. Part of the financial support for the present study came from a grant of the Medical Division of the Rockefeller Foundation to the Institute of Personality Assessment and Research, Berkeley, California.

To say that a remission is spontaneous, however, is generally to make a confession of ignorance; what we mean is that a change has occurred for reasons which we do not understand. When a change in the psychic state of a patient has occurred concurrently with the application of psychotherapy, we usually feel that we *do* understand the causes of the change. Unless, however, we are certain that there is a significant increment in the recovery rate when psychotherapy is applied, we may be deceived in attributing the changes we have observed to the psychotherapeutic forces we think we have applied. Hence the importance of research into psychological change under systematically varied conditions.

The Present Research

This paper reports the results of a study in which 150 psychoneurotic patients all drawn from the same clinic population were tested with the Minnesota Multiphasic Personality Inventory before and after an interval of time during which some of them received psychotherapy and some of them did not. All 150 patients had applied for psychotherapy in the psychiatric clinic in which the study was conducted, and all of them had been accepted for treatment; a group of 23 cases, however, had to be placed on a waiting list until therapeutic facilities became available, and they waited some six months before beginning treatment. Of those who received psychotherapy immediately, 85 received group therapy, and 42 received individual therapy; they included only those patients treated for

a minimum of three months during the period of time covered by the study.

The aim of the study is to discover what changes, if any, occur in MMPI profiles of patients who receive psychotherapy, and to compare these changes with whatever changes may be observed in the MMPI profiles of patients who during the same period of time remain untreated.

Other investigators, such as Kaufmann (5), Gallagher (4), Mosak (7), and Rashkis and Shaskin (8), have reported observations of MMPI changes following psychotherapy; the distinctive feature of the present research is the use of a control group consisting of patients who are in need of therapy and who want it, but who do not receive it.

The psychotherapists who worked with the treatment groups included psychiatrists, social workers, and psychologists. All therapists had at least three years of postdoctoral or postgraduate training and experience. The theoretical orientation in individual therapy was psychoanalytic, although the psychoanalytic method was not used. The explicit therapeutic strategy or time perspective for individual treatment varied from patient to patient, but what occurred generally was brief, ego-oriented psychotherapy on a once-a-week basis. Group therapy differed somewhat in its greater emphasis upon current interpersonal responses; i.e., the interpretations and focusing remarks of the therapist were centered on the interpersonal tactics of the patient in relation to the therapist and the five other patients in the group, the reactions he provoked from other group members, and the meaning of these interactions in terms of the patient's total life situation.

The Research Situation

This study was done in the psychiatric outpatient clinic of the Kaiser Foundation Hospital in Oakland, California, and was conducted with a minimum of interference with the usual clinic routine. No special arrangements were made for the assignment of patients to treatment or nontreatment groups, and neither patients nor psychotherapists considered themselves involved in a research program.

This is perhaps not the ideal design for a study in which an untreated control group is to be used to establish a base line for evaluating the unique contribution of psychotherapy to the amelioration of the psychoneurosis. In a less real world, for example, the research psychologist in a clinic might assign new patients randomly to treatment and nontreatment groups, and he might also set some standard interval as the period of time which should elapse between the determination of the patient's state at the beginning and at the end of the experiment. The experimental (treatment) and control (nontreatment) groups would then be comparable in terms of principle of selection, and might be expected not to differ significantly from one another in severity of initial condition, diagnosis, or prognosis. Furthermore, an equal length of time would have elapsed between $t - 1$ (initial state) and $t - 2$ (terminal state) for both groups.

Under the pressing practical demands of clinical work, it is difficult to arrange a research design which will satisfy perfectly these ideal conditions. The decision not to treat is always a critical one in any therapeutic practice, and in psychiatric work especially such risks as suicide, psychosis, and various other personal and interpersonal catastrophes are possibly attendant upon an unfortunate decision. Random selection of patients for treatment and nontreatment groups is therefore not possible in a situation where the medical responsibility is properly the paramount one.

It is also not possible to establish in advance the exact duration of treatment. Treatment may be terminated at the decision of the patient at any time. Furthermore, early termination of treatment is sometimes indicated for technical reasons, and is brought about by the therapist. Again, in some cases the tempo of work is much slower, and treatment necessarily takes longer. Finally, accidents of circumstance, such as a change in residence or in place of employment, may operate to cause interruption of treatment before the lapse of time specified as desirable for the experiment.

In spite of these irregularities, the design of the present study remains fairly clear-cut.

As it happened, the experimental and control groups turned out to be quite comparable with one another in diagnosis, prognosis, severity of initial condition, age, sex, and education. These comparisons will be presented below. Furthermore, the interval between $t - 1$ and $t - 2$ was an average of 7.00 months for the nontreatment group, 8.32 months for the individual psychotherapy group, and 8.63 months for the group therapy group. These means do not differ significantly from one another (F of 2.5, $p > .05$).

Characteristics of the Sample

None of the patients in this study was so ill at the time of application for treatment as to require hospitalization. They were diagnosed most commonly as psychosomatic neuroses, obsessive or schizoid characters, or phobics (anxiety neuroses, etc.); less commonly, as psychopathic characters or as hysterics. The average age of the group was 32 years, the average education was one year of college. Of the 150 patients, 97 were women; this preponderance of women is characteristic of the clinic.

The difference in proportions of men to women in treatment and nontreatment groups was not statistically significant: women constituted 67 per cent of the nontreatment group and 63 per cent of the treatment group, a percentage difference which gives a CR of 0.37, $p > .05$. No sex differences were found in any of the further comparisons presented in this paper, though computations were done separately for men and women; combined totals for both sexes are therefore given when statistics are cited.

It was indicated above that treatment and nontreatment groups did not differ significantly in age, education, or diagnosis. More exactly, they compared as follows: In age, the mean of the treatment group was 32.48, that of the nontreatment group 30.9; the sigmas were 6.8 and 5.7, respectively. In years of schooling, the mean of the treatment group was 13.6, that of the nontreatment group was 12.8; the respective sigmas were 2.7 and 2.3. In diagnosis, the group percentages in the types of neuroses recognized in the simplified, MMPI-based diagnostic scheme (6) used in the Kaiser Foundation Psychological Research Project were as follows (treatment group percentage given first, nontreatment group next): Psychosomatic 19 per cent, 18 per cent; Hysteric 6 per cent, 9 per cent; Psychopathic 7 per cent, 5 per cent; Obsessive 19 per cent, 23 per cent; Schizoid 21 per cent, 23 per cent; Phobic 12 per cent, 18 per cent; Unclassified 16 per cent, 5 per cent. None of the percentage differences in any category is statistically significant; the largest CR, that for the difference in the unclassified category, was 1.54, the others being less than 1.00.

The patients in the treatment group who were assigned to individual psychotherapy were also compared as a class with those patients assigned to group psychotherapy. These two treatment subgroups did not differ significantly from one another in any of the comparisons presented above. Hereafter the descriptive statistics for the two subgroups will be presented in detail, since it is of some interest to compare the amount and kind of change following group psychotherapy with the amount and kind of change following individual psychotherapy.

The logic of the experiment requires that treatment and nontreatment groups be comparable in terms of prognosis and of severity of initial condition, as well as of diagnosis. The MMPI provides a basis of comparison on both points. The test was scored not only on the commonly used clinical and validity scales, where scores would indicate severity of the psychoneurosis, but also on a new scale developed by one of the authors (1). This scale, called the Ego-strength (Es) scale, was designed originally to predict response of psychoneurotic patients to psychotherapy, and it proved to have substantial validity for that purpose in three samples beyond the standardization sample. Its other correlates in both normal (2) and clinic samples led to its being interpreted as a measure of ego strength, and it was accordingly so named; however, its most important clinical function is as a prognostic indicator.

On all of these MMPI scales, then, there should be no difference between the treatment and the nontreatment groups at $t - 1$. The appropriate comparisons are presented in Table 1.

Table 1

Comparison of Treatment and Nontreatment Groups on MMPI Scales at $t-1$

MMPI scale	Individual psychotherapy ($N=42$)		Group psychotherapy ($N=85$)		No treatment ($N=23$)		
	Mean	*SD*	Mean	*SD*	Mean	*SD*	*F*†
1. Lie	48.04	6.76	47.70	5.25	49.70	5.19	1.03
2. *F*	61.00	12.01	60.14	12.16	58.74	7.94	0.28
3. *K*	51.16	7.29	51.34	8.31	54.96	4.89	2.25
4. *Hs*	61.40	13.61	61.18	13.19	60.04	15.43	0.08
5. *D*	78.71	12.25	73.60	13.72	76.13	12.92	2.08
6. *Hy*	67.21	8.87	66.40	10.38	66.87	13.78	0.08
7. *Pd*	67.57	14.65	66.75	12.08	67.83	12.73	0.09
8. *Pa*	63.66	10.22	61.17	10.76	62.13	8.37	0.82
9. *Pt*	73.09	13.75	69.52	12.70	72.65	12.00	1.29
10. *Sc*	71.50	17.56	68.00	13.69	67.83	12.81	0.88
11. *Ma*	55.88	10.51	54.80	10.32	49.96	10.05	2.84
12. *Es**	41.05	3.46	41.76	7.35	40.80	6.16	0.24

* Raw scores rather than *T* scores are given for *Es*. The mean for clinic populations is 41, and the *SD* 7.5.
† With *df*'s of $n_1 = 2$ and $n_2 = 147$, the *F* ratio must be 3.00 or greater to be significant at the .05 level of confidence. None of the *F* tests shown here indicate statistically significant differences.

As Table 1 shows, the groups did not differ significantly from one another either in initial severity of condition as indicated by the MMPI clinical scales, or in prognosis, as indicated by the *Es* scale. Since they were similar also in diagnosis, and in age and education, it may be concluded, as far as the main variables relevant to this study are concerned, that all three subsamples were drawn from the same population.

Since this is so, the essential conditions on which the groups are varied are these: (*a*) whether or not the patients received psychotherapy in the interval between $t-1$ and $t-2$; and (*b*) whether the patients who received psychotherapy were given individual or group treatment. The relationships between these conditions and the quantity and kind of change, as shown in the differences between MMPI profiles at $t-1$ and $t-2$, will be described next.

Change with Psychotherapy

For each MMPI scale, the discrepancy between the scale score at $t-1$ and at $t-2$ was computed for each patient. If no change at all had been brought about by psychotherapy, one would expect the mean discrepancy scores on all MMPI scales for the treatment groups to vary in a chance manner about zero. That such is not the case may be seen from Table 2.

As Table 2 shows, a significant decrease occurs in the Depression, Hysteria, Hypochondriasis, and Lie scores following psychotherapy, whether individual or group. With group psychotherapy, there is a significant decrease in the Paranoia and Psychasthenia scores as well. With both forms of therapy, there is a significant rise in the Ego-strength scores. Finally, for the individual psychotherapy patients there is also a significant rise in the *K* scale.

These findings accord closely with those reported by other investigators. Gallagher (4) found, in a study of 42 university students treated by the nondirective counseling method, that *D*, *Hs*, *Pt*, and *F* scores decreased significantly following therapy, while *K* increased significantly. Rashkis and Shaskin (8) found significant decreases in *D*, *Hs*, *Hy*, and *Pt* following group therapy. Mosak (7) found decreases in *D*, *Hs*, *Hy*, *Pa*, and *Sc* following individual client-centered counseling in 28 cases. Kaufmann (5) studied 51 university students who had been rated as "improved" following psychiatric treatment, and found significant decreases in *D*, *Hs*, *Pt*, *Sc*, *Pa*, and *F*, and a significant rise in *K*.

At this point a word about the measure of change seems in order. Since the MMPI calls upon the patient to testify concerning himself, and particularly concerning his psychiatric condition, it is subject to all the kinds of distortion to which self-representation is so peculiarly liable. Taken in a clinic, where it is presumably to be scrutinized closely by persons who can be, or already are, very significant in one's life, it is likely to become a communication beyond and apart from the overt meaning of the words. It may be a plea, or a reproach, or a rebuff, or an act of admiration; in brief, an interpersonal communication which expresses something of the relationship of the individual patient to his therapist and to the clinic.

This being inevitably so, it would seem to behoove the psychological researcher to look for evidence of change in other kinds of functions and performances for consideration with the MMPI. In the absence of such other evidence, all that one can do is keep in mind that there is a possibly distorting element of self-representation present in questionnaires. (Perhaps, however, rather too much has been made of the fakeability of tests; life itself can be faked, and many a "cure" is maintained for a time in the interests of an unresolved transference. Furthermore, it may reasonably be argued that durable character traits have a way of showing themselves without one's conscious or even unconscious consent, and that the validity of so well-studied an instrument as the MMPI is not likely to be massively influenced by momentary set.)

Percentage of Improvement

In addition to measuring change simply by considering the average discrepancy score for each scale, it is of some interest to inquire into the *percentage* of change for the better on the scales which do show change. On the scales which most directly reflect neurotic symptoms of which the patient complains (*Hs*, *D*, and *Hy*), the percentage of change in the direction of fewer complaints ranges in the present study from 61 per cent to 67 per cent. This observation concords well with the estimate which Eysenck reports as an over-all rate of improvement with therapy from the data he collated. His judgment was that an estimate of improvement or recovery in two-thirds of the cases would prove "not very far out." The Ego-strength scale, as a matter of fact, showed an increase on second testing in 66 per cent of the cases. None of the other scales changed in the "improved" direction in as many as 60 per cent of the profiles in the sample; the Lie scale, however, did decrease on second testing more than twice as often as it increased (52 per cent vs.

Table 2

Discrepancy Scores ($k_{t_1} - k_{t_2}$) on MMPI Scales for Treatment Groups and Nontreatment Group

MMPI scale	Individual psychotherapy ($N=42$)			Group psychotherapy ($N=85$)			Nontreatment group ($N=23$)		
	Mean	*SD*	*t*	Mean	*SD*	*t*	Mean	*SD*	*t*
1. Lie	1.79	4.46	2.57*	1.58	5.20	2.78*	3.09	4.84	2.99*
2. *F*	0.74	14.67	0.32	0.32	8.07	0.36	−1.13	4.79	1.11
3. *K*	−2.71	7.48	2.32*	−0.86	6.55	1.20	1.30	5.63	1.08
4. *Hs*	3.73	11.29	2.11*	2.50	9.59	2.39*	3.00	8.01	1.76
5. *D*	8.26	15.63	3.38*	2.58	11.60	2.04*	2.74	9.66	1.33
6. *Hy*	4.43	10.64	2.66*	2.51	8.23	2.79*	3.48	9.37	1.74
7. *Pd*	0.57	12.45	0.29	1.19	9.19	1.19	4.57	9.47	2.26*
8. *Pa*	1.69	10.58	1.02	2.09	6.99	2.74*	1.91	8.25	1.08
9. *Pt*	4.00	14.44	1.77	2.58	11.54	2.05*	1.04	8.56	0.57
10. *Sc*	1.38	16.52	0.53	1.01	10.80	0.86	−2.13	8.61	1.16
11. *Ma*	−0.52	11.26	0.30	0.22	8.71	0.23	−1.61	8.68	0.87
12. *Es*†	−2.62	6.60	2.38*	−1.40	7.14	2.10*	−1.80	3.85	2.04*

* Mean discrepancy significantly different from zero at .05 level of confidence.
† Computations based on raw scores rather than *T*-scores.

25 per cent), but also remained the same much more often than did the other scales.

The immediate general conclusion to be drawn from these data is that the psychotherapy patients, if one is to judge from their MMPI performances, are less ill than when they began treatment. The observed changes are primarily of a symptomatic rather than of a characterological sort: the patients are in a better mood (less depressed), are not as hypochondriacal, and are both less defensive (as shown in the Lie scores) and less given to repression as a defense mechanism (the Hysteria scale). The increase in the *K* score in the individual psychotherapy patients no doubt merits special comment, especially since it was found to increase following therapy in some of the other studies cited. The common interpretation of *K* is that it indicates defensiveness, and that a high *K* score betokens a test-taking attitude which involves representing oneself in a better light than actual clinical condition would warrant. However, in studies over the past five years at the Institute of Personality Assessment and Research, *K* has regularly been observed to correlate positively with favorable aspects of personal functioning. In the present study, it should be noted that while *K* is rising, *L* is falling; if indeed an attempt were being made by these patients to represent themselves in an unwarrantedly better light at second testing, one would expect *L* to rise as well. Perhaps the *K* scale needs a systematic re-evaluation before being accepted as a measure of test-taking attitude.

Changes in the Control Group

Patients who received psychotherapy clearly responded differently to the MMPI when they were tested at the conclusion of the therapy; what, then, of the patients who remained untreated for a comparable length of time? Does time alone have an equal effect, or does treatment make a significant difference? A partial answer is provided by the comparable statistics for the nontreatment group, which are presented in Table 2.

From Table 2 it is evident that the patients who were waiting for treatment had somewhat fewer complaints on second testing, although most of the clinical scales do not change significantly. On the neurotic triad (*D, Hs, Hy*) the average decrease is slightly greater for the controls than for the group psychotherapy patients. The percentage of change for the better in those scales averaged 61 per cent in the control group, which is slightly, but not significantly, less than in the psychotherapy groups. There is a significant increase in the control group in scores on *Es*, which would suggest that at the end of the waiting period the patients are somewhat more amenable to psychotherapy than at the beginning of that period. The actual percentage of patients who improved on *Es* in the control group was 55 per cent, compared with 66 per cent in the therapy groups, a difference which is not statistically significant.

One might question at this point whether the patients on the waiting list should really be thought of as "untreated." In a sense, of course, simply having committed oneself to participating in psychotherapy, and having had a reciprocal commitment from a clinic to afford psychotherapy, even though not immediately, represents a breaking of the neurotic circle. A force for change has already been introduced. In addition, the initial interview and the psychological testing may themselves be therapeutic events, since during such sessions the patient makes some efforts to confront himself and his problems more objectively than he has in the past. Moreover, the focus of the intake interview at the Kaiser Foundation clinic was generally upon realistic questions concerning the patient's situation and possible courses of action, so that even the waiting-list "controls" might well feel that a realistic review of their situation had been achieved, that an important decision to change had been made by them, and that in general they had obtained more control over their own lives. As the statistics in Table 2 indicate, these changes are slight and are not statistically significant for the clinical scales (with the exception of *Pd*), but there is *some* change. From a methodological standpoint, it may be suggested that other control groups should be added in future studies, since the psychological factors involved in being on a waiting list are themselves possibly determinative of changes in psychiatric condition.

Table 3

Comparison of Treatment and Nontreatment Groups in Terms of Discrepancy Scores ($k_{t_1} - k_{t_2}$) on MMPI Scales

MMPI Scale	Individual psychotherapy vs. nontreatment		Group psychotherapy vs. nontreatment	
	Diff. between mean discrepancy scores	*CR*	Diff. between mean discrepancy scores	*CR*
1. *L*	1.30	1.06	1.51	1.31
2. *F*	1.87	0.75	1.45	1.10
3. *K*	4.01	2.44*	2.16	1.58
4. *Hs*	0.73	0.30	0.50	0.25
5. *D*	5.52	1.75	0.16	0.06
6. *Hy*	0.95	1.49	0.97	0.45
7. *Pd*	4.00	1.41	2.38	1.08
8. *Pa*	0.22	1.25	0.18	0.15
9. *Pt*	2.96	1.11	1.54	0.70
10. *Sc*	3.51	0.68	3.14	1.47
11. *Ma*	1.09	0.44	1.83	0.89
12. *Es*	0.82	0.64	0.40	0.36

* Difference statistically significant at .05 level.

The question remains, is there significantly more change with regular and formally defined psychotherapy than without it? The answer is provided in Table 3, which presents comparisons of the treatment and nontreatment groups in terms of discrepancy scores. When the amount of discrepancy consequent upon nontreatment is subtracted from the amount consequent upon either individual psychotherapy or group psychotherapy, what remains is rarely a discrepancy of such magnitude as to be statistically significant. The only comparison which did yield a significant difference in discrepancy scores was the comparison of the individual psychotherapy group with the nontreatment group on the *K* scale; and, in that instance, the difference was large because in the case of the nontreatment group the discrepancy was positive (decrease in *K* scores), while in the individual psychotherapy group the discrepancy was negative (rise in *K*). For the most part, however, the changes tended to be in the same direction for treatment and nontreatment groups, and of almost equal magnitude.

Summary

The results yielded by this study may be summarized as follows:

1. The mean pre-post MMPI scores of patients undergoing both group and individual therapy showed significant improvement on symptom scales.

2. Patients who decided to enter therapy but who were required to wait for treatment showed some MMPI improvement during the waiting period, primarily in prognosis.

3. The therapy patients did not improve significantly more than did the waiting-list controls.

4. These results suggest that the use of "waiting-list" controls in investigation of changes in therapy may be complicated by therapeutic factors involved in the initial intake evaluations and decisions.

5. The use of additional "nonwaiting-list controls" is suggested for future investigations on the effect of psychotherapy.

Received November 3, 1954.

References

1. Barron, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 327–333.
2. Barron, F. Ego-strength in normal men as measured by an MMPI scale. Paper read at the West. Psychol. Ass., 1954. Library of the Institute of Personality Assessment and Research, Univer. of California, Berkeley.
3. Eysenck, H. J. The effects of psychotherapy: an evaluation. *J. consult. Psychol.*, 1952, **16**, 319–324.
4. Gallagher, J. J. MMPI changes and client-centered therapy. *J. consult. Psychol.*, 1953, **17**, 334–338.
5. Kaufmann, P. Changes in the Minnesota Multiphasic Personality Inventory as a function of psychiatric therapy. *J. consult. Psychol.*, 1950, **14**, 458–464.
6. Leary, T., & Coffey, H. The prediction of interpersonal behavior in group psychotherapy. *Group Psychother.*, 1954, **7**, 7–51.
7. Mosak, H. H. Evaluation of psychotherapy: a study of some current measures. Unpublished doctor's dissertation, Univer. of Chicago, 1950.
8. Rashkis, H. A., & Shaskin, D. A. The effects of group therapy on personality scores. *Amer. J. Orthopsychiat.*, 1946, **16**, 343–349.
9. Rosenzweig, S. A transvaluation of psychotherapy —a reply to Hans Eysenck. *J. abnorm. soc. Psychol.*, 1954, **49**, 298–304.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Brief Psychotherapy and Enuresis[1]

Daniel H. Harris, Richard W. Firestone, and Carl M. Wagner
Marine Corps Recruit Depot, San Diego, California

The present study derives from the premise that nocturnal enuresis in adults is primarily psychogenic in origin. Supporting this was the observation that the psychological climate induced by military training could often be meaningfully related to variations in the enuretic symptoms of Marine recruits. This experiment investigated the effect of a psychotherapeutic situation created prior to training on the adjustment of such recruits. The specific hypothesis tested was: a significant reduction of admissions to the psychiatric observation ward in which enuresis is a factor will result if, during the regular screening procedure, enuretics receive brief psychotherapy combining recognition and acceptance of their problem with specific suggestions for overcoming the habit.

Two hundred recruits (ages 17 to 27) entering training with an admitted history of enuresis during the previous five years were randomly assigned to an experimental group ($N = 100$) or control group ($N = 100$). The two groups did not differ significantly in age, education, intelligence, or degree of apparent neuropsychiatric involvement.

For experimental subjects, the regular psychiatric screening interview was expanded into a brief psychotherapeutic session designed to relieve anxiety related to enuresis by means of a permissive and reassuring discussion of the symptom, plus a focusing of positive motivation around some plausible behavioral suggestions for overcoming the habit. *Ss* in the control group had only the regular screening interview. Subsequent contacts with *Ss* in either group followed a standardized pattern in which further attempts at therapy were avoided.

On the basis of a rigid set of admission criteria, 20 control *Ss* were admitted to the psychiatric observation ward prior to completion of recruit training, while only eight were admitted from the experimental group. Chi-square analysis revealed this difference to be significant, supporting the hypothesis at better than the .02 level of confidence. Apparently the therapeutic situation utilized in this study was instrumental in substantially reducing the number of ward admissions.

Among a third group of recruits who concealed histories of enuresis at screening but later proved to be enuretic, the ward admission rate was five times as high as that for experimental and control groups combined, the difference being significant at the .01 level. This finding gives additional support to the main hypothesis.

Brief Report
Received April 19, 1955.

[1] Opinions or conclusions contained in this paper are those of the authors. They are not to be construed as necessarily reflecting the view or endorsement of the Navy Department.

An extended report of this study may be obtained without charge from Daniel H. Harris, Marine Corps Depot, San Diego, California, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4592 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

The Measurement of Depth of Interpretation[1]

Norman I. Harway,[2] Allen T. Dittmann,[3] Harold L. Raush,
Edward S. Bordin, and David Rigler
University of Michigan

When the proponents of one viewpoint of psychotherapy criticize the studies of any other school of thought, the criticisms tend to fall in one or both of two classes: (*a*) the concepts and relationships studied are unimportant or have only minor relevancy; (*b*) the conditions of measurement are such as to preclude proper investigation of the variables under consideration.

The study reported here is one of a series of investigations of psychotherapeutic processes (1). In these studies we have attempted to select variables which have theoretical and practical relevance over a wide variety of therapeutic approaches. Furthermore, we have concerned ourselves intensively with issues of measurement. In these ways our eventual goals of hypothesis testing can, we hope, communicate beyond the narrow barriers of school partisanship.

One of the major activities of the psychotherapist in the course of treating his patient is to engage or avoid engaging in interpretive behavior (5, 9). Collier has presented a rather full discussion of the interpretive continuum as it relates to psychoanalytic and nondirective therapy (2). Like Collier's study (3), the present paper reports the development of a scale for the measurement of a therapist's depth of interpretation. We felt, however, that the immediate utilization of such scales for hypothesis testing would be premature. In studies of psychotherapy, as in all studies involving ratings and codings, interpretation of data is limited by our understanding of the conditions of data collection. So long as these conditions are not understood and accounted for, investigators will arrive at discrepant results, and will question one another's data.

Typical among measurement issues raised with regard to variables in psychotherapy are questions concerning context, unit size, and method of data presentation. Each of these issues becomes the focus for sometimes naive, sometimes sophisticated criticism of published studies. For example, we may consider the kinds of statements that are apt to be made in discussions of research in psychotherapy: "The failure to provide judges with information as to what has gone on previously in therapy, makes their judgments of the behavior in question meaningless" (the problem of context); or, "What is *really* meant by the concept is not any specific, momentary behavior, but rather the over-all effect" (the problem of unit size); or, "The judges would have to have movie films (or recordings) in order to *really* evaluate the behavior" (the problem of mode of presentation).

The specific definition of depth of interpretation used in the present series of studies is: "Any behavior on the part of the therapist that is an expression of his view of the patient's emotions and motivations—either wholly or in part—is considered an interpretation. A patient has varying degrees of awareness of his emotions and motivations. Depth of interpretation is a description of the relationship between the view expressed by the

[1] This paper is part of an investigation supported by a research grant M–516 C–2 from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service. The authors are indebted to Dr. R. L. Cutler, Dr. W. L. Hays, Mr. Z. Sperber, and Miss J. V. Williams for their assistance in various aspects of the research.

[2] Now at the University of Rochester, Rochester, New York.

[3] Now at the National Institute of Mental Health, Bethesda 14, Maryland.

therapist and the patient's awareness. The greater the disparity between the view expressed by the therapist and the patient's own awareness of these emotions and motivations the deeper the interpretation." The definition is compatible with both psychoanalytic and nondirective viewpoints as expressed in the literature (5, 9) and is sufficiently broad to encompass the data of a variety of therapeutic approaches.

Acceptance of the definition of interpretation as an interaction of patient and therapist has certain implications for the measurement operations discussed above:

(*a*) *With regard to unit size:* Since we could see no firm basis for an a priori choice as to unit, we felt that a comparison of extremely divergent sizes would best indicate any effects that might go with variation in size.

(*b*) *With regard to method of presentation:* Since the variable of depth of interpretation seems in the main to deal with the relation between the content of the therapist's behavior and the content of the patient's awareness, we should expect little or no difference in the evaluations obtained by using the typescript of an interview as compared with using the actual sound recording.

(*c*) *With regard to context:* Clearly the interpretive value of a therapist's statements is not determined *in vacuo* but in relation to the determination of what the patient "actually" means and how conscious or aware he is of this actual meaning. Thus, if we force judges to rate depth of interpretation without giving them information on what the patient has said or is saying, the ratings should have little or no correlation with ratings obtained when the necessary information is available.

Development of Rating Scale

Following Thurstone's method of equal-appearing intervals (10), a seven-point scale for rating depth of interpretation was developed. Seventy statements descriptive of therapist activity were sorted by two groups of judges, one experienced, the other naive. The experienced group consisted of fourteen staff members and advanced graduate students in clinical psychology, each with at least 100 hours of experience in psychotherapeutic practice. Eleven undergraduate students in the first course in psychology composed the "naive" judges. Judges with disparate experience were used in hope of identifying items highly communicable to a wide range of possible raters.

The importance of taking judge groups into account was demonstrated by the fact that the naive judges were significantly more variable and placed significantly more items at the deeper end of the scale than did the experienced judges.

Seven of the statements had been originally selected so as not to fit the requirements of our definition; while descriptive of therapist activity they did not give any indication of patient awareness or of the relation of the therapist's activity to the patient's awareness, e.g., "Questioning designed to gain information," "Therapist suggests that the patient desires to terminate therapy." These seven statements offered an internal check. If they scaled along with our bona fide statements, the meaning of the resulting scale would be open to question. None of these "spurious" statements met our criteria for inclusion in the scale.

A seven-point graphic scale was obtained. Nine statements were selected to exemplify points on this scale. The statements and their scale positions (running from low to high depth) are:

(*a*) Therapist merely repeats the material of which the patient is fully aware (1.1).

(*b*) Restatement of material of which the patient is aware (1.8).

(*c*) Implied focusing with regard to material of which the patient is aware (2.3).

(*d*) Therapist connects for the patient two aspects of the content of the previous patient statement (3.4).

(*e*) Reformulation of the behavior of the patient during the interview in a way not explicitly recognized previously by patient (3.9).

(*f*) Therapist comments on patient's bodily and facial expressions as manifestations of patient's feelings (4.4).

(*g*) Use of preceding patient statement to exemplify a process that has been building up during an interview and of which the patient is seemingly unaware (4.9).

(*h*) Therapist speculates as to a possible childhood situation that might relate to current patient feeling (5.9).

(*i*) Therapist response deals with inferences about material completely removed from the patient's awareness (6.9).

In the studies below the definition of depth of interpretation and the scale were presented with the statement that the nine items on the scale "are not inclusive of all possible descriptions of depth at these points . . . but are merely examples of types of things that could occur and where they would fall on the scale."

The Conditions of Measurement Study

Design

Sixteen raters used the scale to evaluate four therapy interviews, each taken from a different case and each involving a different therapist. The raters had the same qualifications as those of the experienced judges involved in the development of the scale. The three sets of measurement conditions discussed earlier were investigated. These were: (*a*) the unit of the interview to be rated—the individual therapist response vs. the interview as a whole, (*b*) method of presentation of the interview to the rater—typescript vs. tape recording, (*c*) amount of information or context given to the rater—the interview to be rated with its preceding interview (preceding interview condition) vs. only the therapist's responses in sequence from the interview to be rated (therapist only condition).

The study was designed to control systematically or evaluate the influence of raters, interviews and practice in the use of the scale. The time-consuming nature of the task did not allow complete control of these factors. A relatively great amount of time is necessary to read or to listen to a number of 50- or 60-minute interviews. Also to be considered were the number of raters available with appropriate experience and training and the number of clearly recorded psychotherapy interviews obtainable.

A 2× 2 × 2 factorial design was projected (each of the three variables—method of presentation, context, and unit size—was varied over two levels). To form a latin square such that each rater rated under each combination of the experimental conditions in a different order than any other rater, and such that each combination of the experimental conditions was rated but once at each order position from one to eight, required eight raters. Table 1 presents the eight possible combinations of the three factors and the assignment of the raters to the various combinations. The design was replicated, giving us a total of sixteen raters. Kogan (7) has noted the advantages in certain instances of "confounding" in the design of psychological experiments. In the present instance, confounding of the three-factor interaction (Presentation × Context × Unit) with the rater differences reduced the interviews to four and the time required by half.[4]

Table 1

Assignment of Raters to Experimental Conditions

	Typescript		Recording	
Context	Interview unit	Response unit	Interview unit	Response unit
Therapist only	I*	V	V	I
	II	VI	VI	II
	III	VII	VII	III
	IV	VIII	VIII	IV
Preceding interview	V	I	I	V
	VI	II	II	VI
	VII	III	III	VII
	VIII	IV	IV	VIII

* Roman numerals represent raters.

When the interview was the unit rated, only one score was obtained from the rater. When the response was the unit rated, there were, of course, as many scores for each rater as there were responses. In the latter case the response unit ratings for a single judge were averaged to give a rating for the interview as a whole, and then mean response unit ratings were compared with ratings of interviews as units. Regardless of the method by which

[4] Among the characteristics of these interviews: The settings were a University counseling center, a Mental Hygiene Clinic, an out patient clinic and a NP Hospital. The hours represented were the 20th, 40th, 146th and 11th. Therapists were of analytic or modified analytic orientation. They ranged in experience from less than 1 year to 10 years of therapeutic work.

these ratings were obtained, they are assumed to represent measurements of one aspect (i.e., depth of interpretation) of the same fundamental variate (behavior). Although some of the entries in this analysis were arrived at by averaging a set of ratings of responses, these averages are not to be interpreted in the usual sense of representing a number of independent observations of the fundamental variate. They are, instead, meant to represent only ratings of behavior arrived at in a particular way—the rating of a whole by means of ratings of component parts.

Another problem in connection with the use of response units was the pertinence of depth of interpretation to the various responses in an interview. This was resolved by instructing raters to rate each unit but to indicate those units or responses which they did not feel were ratable on this particular scale. The ratings to responses so indicated were considered as "forced" ratings and were not used in computing the means for the response unit scores nor were they used in computing interrater agreements in rating.

Results

As seen in Table 2, the scale of depth of interpretation significantly differentiated among the interviews at the .01 level. In fact, the variance for interviews was about twice that attributable to pooled experimental conditions. The experimental conditions were found to be a significant source of variance at the .01 level. Any interview by experimental condition interaction, if such existed, could not be evaluated.

Variation due to sequences was not significant. The difference between raters within a sequence was significant at the .05 level. Such a finding could result from the use by raters of different portions of the scale. It need not affect the agreement between raters in the ordering of responses or interviews on the scale.

Unit size. Since each experimental condition represents a specific combination of three

Table 2

Analysis of Variance

Source of variation	Sum of squares	*df*	Mean square	*F*
Sequences	5.69	7	.81	1.89
Raters within sequences	8.50	8	1.06	2.47*
Total between raters	14.19	15		
Order	.18	3	.06	—
Interviews	12.64	3	4.21	9.78**
Experimental conditions	15.23	6	2.54	5.89**
Presentation	.30	1	.30	—
Context	.49	1	.49	1.14
Units	13.32	1	13.32	30.93**
Presentation × context	.42	1	.42	—
Presentation × units	.49	1	.49	1.14
Context × units	.20	1	.20	—
Pooled error	15.51	36	.43	
Greco-Latin square error	4.61	12	.38	
Pooled rater × order interaction, different sequences	10.90	24	.45	
Total within raters	43.56	48		
Grand Total	57.75	63		

* Significant at .05 level.
** Significant at .01 level.

factors, the sum of the squares and degrees of freedom could be broken down into the component parts. Testing each of the component mean squares for the experimental conditions against the pooled error mean square indicates that the difference in ratings due to experimental conditions was primarily the result of one factor, namely, the size of the unit evaluated. For each interview the evaluation based on the interview unit was "deeper" than the mean of the response unit ratings. A further study showed that this result was not a function of the exclusion of forced ratings. Apparently the judge's perception of the depth of an entire interview is strongly determined by the therapist's deeper responses. This finding is of particular significance because previously many investigators have made the assumption that the rating of an entire interview was equivalent to the pooling of a sequence of response ratings within an interview in terms of some average rating. As Mowrer, Hunt, and Kogan note in a slightly different context, "the fact that a client mentions one problem ten times and a second problem just once does not necessarily mean to the clinician that the first problem is ten times as important to the client" (8, p. 259).

Method of presentation. It had been expected that ratings of depth of interpretation would not differ when based on a typescript as compared to a tape recording of a therapy interview. The basis for this expectation was the fact that this variable was primarily a "content" variable, that is, that its rating depended upon the actual words spoken, and not upon some other factor such as tone of voice, facial expression, etc. The analysis in Table 2 indicates that there is no difference in the ratings of interviews under different methods of presentation of the interview.[5] For evaluations of this particular variable, then, we can in the future present typescripts to our raters and enjoy the practical advantages stemming from the easier administration permitted by this method.

[5] Studies of the effects of inclusion of forced ratings, and studies of the reliabilities of each of the methods and the cross correlations between methods lend further support to the conclusion that evaluations of depth of interpretation do not vary as a function of these alternative methods of presentation.

Context. We turn now to the critical question of context. The concept of interpretation, and the specific definition used in this study imply that knowledge of both patient and therapist activity is necessary for a correct appraisal of the interpretive level. If our measurements—in this instance ratings—are valid, then the removal of the patient responses should lead to different results from those obtained when the patient responses are included in the information given to the raters. The analysis of variance presented in Table 2 indicates that the expected difference did not occur. Furthermore the reliabilities of the two methods are comparable, and the cross correlations indicate that the methods are equivalent.

As has been noted, such a finding casts doubt on the validity of our measurements or on the adequacy of the definition of the concept. A third alternative, however, suggested itself. It seemed possible that the raters were obtaining information as to the patient's behavior from the sequence of the therapist's responses. To check on this, a study was done in which four new raters and two of the interviews were used.

In this study all ratings were made on the basis of typescript and with response as the unit. Each rater evaluated one interview with only the therapist responses in sequence present (this is the therapist only context of the earlier study) and the other interview with only the therapist responses present but in *random* order. The order of presentation of interviews and of experimental conditions was systematically varied.

There is no difference in the mean response unit ratings between ratings based on sequential therapist only information and ratings based on randomized therapist only information. However, the analysis of interrater reliability coefficients and the cross-context correlations, suggests that marked differences do exist between the sequential context and the random order context. The reliability coefficients for sequential order were .64 and .51, while those for random order were .49 and − .03. On the other hand, none of the eight cross-context correlations differ significantly from zero, ranging from − .19 to .15 with a median of − .05. In other words, while raters

might agree within the random order condition, there was no evidence of significant agreement across conditions. These data suggest that the sequence did give the judges enough information about what the patient's behavior may have been to allow them to make ratings equivalent to those of judges who actually had read the intervening patient response. Randomizing the sequence of therapist responses probably randomized the judge's hypotheses about the level of awareness demonstrated by the intervening patient behavior. However, this conclusion was based on a small number of correlations, and one of the two correlations under random order condition was almost as high as those observed under sequential order. Therefore two more judges were asked to rate the therapist responses in this interview under the same two conditions. The four additional sequential order correlations ranged from + .81 to + .52, confirming our previous finding. The four random order correlations were − .03, − .06, − .08 and + .24. The ten additional cross correlations ranged about zero, some, however, reaching the level of statistical significance. These additional data allow us to state with increased confidence that judges were attempting to judge therapist's depth of interpretation as it was defined.

In a further effort to generalize our findings the basic four-judge design was applied to two new interviews. We were surprised to find that the sequential order, random order and cross correlation were all at the same level, median correlations being + .57, + .61, and + .53, respectively. Our present impression is that the therapist style in these interviews (early counseling interviews) will explain the result. Therapists sometimes recapitulate to the patient what he was saying or doing before presenting an interpretive reaction to the patient's activity. Thus, information would be available to a judge no matter where in the sequence the therapist response was located. As a first check on this impression a count was made of the number of words used by therapists in the two studies. As expected, therapists in the interviews used in the latter study tended to make longer responses. The information-bearing characteristics of therapists' communications are now being analyzed and further sequence studies are planned to test these ideas.

Reliability and validity. It might be argued that the failure to find differences in the main study due to context is due to the low reliability of our measure. One of the usual methods for determining reliability of ratings is by means of agreement between raters. Almost all the reliability coefficients based on the ratings of two raters working on the same interview under the same context and presentation conditions were statistically significant. The median coefficient in the main study was .40. Reliabilities of this order are disappointing and at first glance an obvious suggestion would be to dispense with the scale. It must be considered, however, that what is represented by these numerical values is analogous in test development to item reliability. For example, we computed the reliability of response unit ratings of depth for one of our interviews on the basis of a ten-response segment. The means of the ratings for the first ten responses, second ten responses, and so on, were obtained for two raters, and these means were then correlated. Interjudge agreement was now .74 compared to .36 for single response units. As in a test, we may increase our reliabilities by increasing the number of items, each of which may have rather low reliability. The problem, however, remains: Can we study single responses, rather than summing or averaging over responses? Here too, we may make use of test theory. Where one is interested in individual items, where any one judgment may be of low reliability, as in opinion surveys, accuracy of measurement can be attained through increased sampling. We have, for example, found that when we pool the ratings of a number of judges on individual items, we obtain almost perfect correspondence (rhos of .89, .94, and .93) with the ordering arrived at by combined judgments of other judges using other methods, such as rankings and method of similarities by triads (4).[6]

[6] Further confirmation of the usefulness of the scale appears when reliability of individual response items is estimated by the Horst formula (6). Analysis indicates that agreement among the judges accounted for from 39 to 62 per cent of the total variance among items.

Considering the statistical significance of most of the coefficients, the relatively inferential and complex nature of the variable being rated and the fact that none of the raters had any training in the specific use of the scale, the extent of interrater agreement for depth of interpretation is far from disappointing. The certainty with which the four interviews could be discriminated gives substance to this conclusion. Further, if our measurements were sufficiently reliable to differentiate interviews and unit size, it follows that our failure to find differences between the preceding interview and therapist-only context conditions cannot be attributed solely to errors in measurement.

Having considered reliability, we may now summarize our findings concerning the validity of the scale. The later studies indicated that the initial test of validity, based upon a comparison of preceding interview with therapist-only context conditions, was not critical. Further, these later results do provide evidence for validity. Frequently, in the measurement of variables such as the one presented here, validity is not evaluated until evidence is available of success in predicting relationships with other variables. We have illustrated in this paper that it is possible to check on validity of measurement by examining the consistency of the measurements with the definition of the variable.

Summary

In this paper we have illustrated not only how different procedures of measurement lead to differing perceptions of the therapeutic process, but also how varying the measurement conditions can, in certain instances, enable a check on the internal validity of the concept under investigation. Three of the conditions of measurement—method of presentation, unit size, and context—were varied in presenting psychotherapy interviews to raters. Ratings were of the depth of a therapist's interpretations to the patient and were based on a seven-point scale. The four interviews used were clearly discriminated. Variations in unit size and in certain conditions of context led to changes in ratings of depth.

Received October 22, 1954.

References

1. Bordin, E. S., Cutler, R. L., Dittmann, A. T., Harway, N. I., Raush, H. L., & Rigler, D. Measurement problems in process research on psychotherapy. *J. consult. Psychol.*, 1954, **18**, 79–82.
2. Collier, R. M. A basis for integration rather than fragmentation in psychotherapy. *J. consult. Psychol.*, 1950, **14**, 199–205.
3. Collier, R. M. A scale for rating responses of the psychotherapist. *J. consult. Psychol.*, 1953, **17**, 321–326.
4. Coombs, C. H. Theory and methods of social measurement. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences*. New York: Dryden Press, 1953. Pp. 471–535.
5. Fenichel, O. *Problems of psychoanalytic technique*. Albany, New York: Psychoanalytic Quarterly, 1941.
6. Horst, P. A generalized expression for the reliability of measures. *Psychometrika*, 1949, **14**, 21–31.
7. Kogan, L. S. Variance designs in psychological research. *Psychol. Bull.*, 1953, **50**, 1–40.
8. Mowrer, O. H., Hunt, J. McV., & Kogan, L. S. Further studies utilizing the discomfort-relief quotient. In O. H. Mowrer (Ed.), *Psychotherapy theory and research*. New York: Ronald, 1953. Pp. 257–295.
9. Rogers, C. R. *Counseling and psychotherapy*. Boston: Houghton Mifflin, 1942.
10. Thurstone, L. L. Attitudes can be measured. *Amer. J. Sociol.*, 1928, **33**, 529–554.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Responses to Imagined Auditory Stimuli as Compared to Recorded Sounds[1]

David R. Stone
Utah State Agricultural College

The use of auditory stimuli as a source of useful data is in the exploratory stage. Preliminary work has indicated that such an approach can produce useful diagnostic signs, both for clinical and developmental analysis.

The work to date has centered around two major trends, (*a*) presenting sounds in single order, requiring free associative response to the sounds in turn (1), and (*b*) presenting the sounds in sets, asking for multiple association to the sounds taken as a whole (2).

The present study followed the latter trend, and was an amplification of the multiple-association technique as applied to *imagined* sounds, rather than to actual recorded sounds. The sounds were described verbally, in contrast to being played.

In the recorded version, the subjects (220 college students) heard a bird chirping, a train coming from a distance which comes closer until a crashing is heard, and a trumpet fanfare.

The subjects for the study of imagined sounds were 74 students in a general psychology class. The instructions were:

> In a moment I will describe three sounds. When you hear the descriptions, write a story in which you use the sounds as described. Tell what has led up to the sounds, what is happening, and how it turns out. You will have five minutes. For the first sound, imagine that you hear a bird chirping. For the second sound, imagine that you hear a train puffing in the distance. You hear the faint sound of a whistle. Gradually it gets closer and closer. Then you hear a terrific crash. For the third sound, imagine that you hear a trumpet fanfare.

At the end of four minutes, the subjects were told of the time and advised that they would have one minute to complete their story by telling how it turned out.

In terms of length, order of sounds, situations described, characters, sound descriptions, and outcomes, the responses to imagined sounds were either similar to the recorded version, or included them plus additional elaboration. The use of imaginary (verbally described) sounds seems as productive, though less structured than the recorded stimuli.

Perhaps a more subtle analysis would reveal differences not now apparent. It is presumed at present that the association to multiple stimuli is a more crucial element of these responses than whether the auditory stimuli are imagined or real.

Brief Report
Received April 25, 1955.

[1] An extended report of this study may be obtained without charge from David R. Stone, Utah State Agricultural College, Logan, Utah, or for a fee from the American Documentation Institute. To obtain it from the latter source, Order Document No. 4591 from ADI Auxiliary Publication Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm, or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

References

1. Wilmer, H. A., & Husni, May. The use of sounds in a projective test. *J. consult. Psychol.*, 1953, 17, 377–383.
2. Stone, D. R. *The auditory apperception test.* Beverly Hills, Calif.: Western Psychological Services, 1953.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Clinical Statements as Scientific Propositions and Social Decisions[1,2]

Frank M. du Mas
Montana State University

"Tom is a neurotic." "Dick is to remain at home under the continuing care of his parents until further study of his case can be made." "Harry is to be committed to the state institution for mental defectives." These statements, though overly simplified and out of context, are the kinds of statements clinical psychologists must make in their routine professional work. Each statement is different from every other statement in certain respects. The statement about Tom is a more or less straightforward statement of a matter of fact. The statement about Dick is a social decision to withhold positive action. The statement about Harry is a social decision to initiate a definite course of action. The clinical psychologist is an applied scientist and these statements are relevant to the *applied* part of his work. He is a scientist operating within the confines of his social matrix and he is constrained to obey administrative protocol, the laws of his society, and to make decisions about a patient which may radically change the patient's way of life. It is not a question of whether or not a clinician wants to make these kinds of decisions. He must, if he is to work effectively in a social institution such as a hospital.

But the clinician is also a scientist. In his endeavor to push back the frontiers of knowledge, he frequently resorts to theoretical analysis in his study and construction of theoretical systems which organize and integrate his field. When he does this, he turns to deductive logic in order to facilitate his theoretical work. He often restates an inductive generalization as a postulate in a deductive system. Then, as in all theoretical work, he operates within a formal system and the rules of the game are not quite the same as the rules appropriate to inductive empirical science or the rules of society.

It seems, therefore, that the clinician is concerned with statements in three domains: formal, empirical, societal. What is the appropriate logical frame of reference to use in evaluating the validity of statements in each of these domains? Since, as we shall show, eminent authorities give quite different answers to this question, let us refine the issues.

One widely held point of view is well stated by Ayer:

> Like Hume, I divide all genuine propositions into two classes: those which, in his terminology, concern "relations of ideas," and those which concern "matters of fact." The former class comprises the *a priori* propositions of logic and pure mathematics, and these I allow to be necessary and certain. . . . That is, I maintain that the reason why these propositions cannot be confuted in experience is that they do not make any assertion about the empirical world, but simply record our determination to use symbols in a certain fashion. Propositions concerning matters of fact, on the other hand, I hold to be hypotheses, which can be probable but never certain (3, p. 31).

Stated differently, what Ayer is saying is that *formal* propositions may be evaluated by reference to a 2-value logic, that is, any formal proposition is either true or false. In probability terms, the probability of any for-

[1] This article is a major revision of a paper read at the 1954 Convention of the Midwestern Psychological Association.

[2] The ideas presented here are relevant to all applied sciences. The discussion is pointed at one particular field, clinical psychology, in order to simplify the discussion. An understanding of our ideas in one applied science should easily transfer to other applied sciences.

mal proposition being true is either 0 or 1. *Empirical* propositions, however, can never be certainly true or certainly false—they can only be probable. That is, an empirical proposition has a truth value which may be represented by a point in a truth continuum. In probability terms, the probability of an empirical proposition being true is always greater than 0 but less than 1. The truth (or probability) continuum implies that a multivalue logic is necessary for the evaluation of empirical propositions because a continuum contains an infinite number of possible points or truth values.

Another point of view maintains that it is possible to evaluate a proposition about an object, person, or event as either certainly true or certainly false. This point of view is implicit in the writings of Allport:

> The truth of the matter seems to be that *this* delinquent has either 100 per cent certainty of becoming a repeater or 100 per cent certainty of going straight (2, p. 156).

In another source Allport states:

> Indeed, if we believe in determinism at all, his (the parolee) chances are either zero or 100: he is bound to succeed or else to violate . . . and (ideography) will bring your prediction . . . to a 100% certain individual prediction (1, p. 17).

Others agree with Allport. Snygg and Combs state:

> Far from being inconsistent, behavior becomes characterized by complete predictability (4, p. 58).

Snygg and Combs (4, p. 437) affirm their agreement with Allport and include a quotation from his work.

Stated differently, what these writers are saying is that *empirical* propositions may be evaluated by reference to a 2-value logic, that is, that it is at least sometimes possible to say that an empirical proposition is certainly true or certainly false. In terms of probability, they say the probability of some empirical propositions is either 0 or 1.

These representative quotations raise certain questions which need to be answered. What logical frame of reference is appropriate to formal statements? To empirical statements? To societal statements? How many and what kinds of logical frames of reference are required by the clinician? Can a unitary, systematic integration be made of the several logical frames of reference?

Let us attempt to answer these questions.

Table 1

Logical Frames of Reference and Their Applications to Specific Clinical Statements

Relations	Truth values*	Applications
I	$0 \leq p \leq 1$	All clinical statements
II	(F) (T)	Formal propositions
III	$p = 0, 1$	Formal propositions
IV	$0 < p < 1$	Empirical propositions
V	(F) (D) (T)	Subjective social decision: withholding action
VI	$0 < F \mid D \mid T < 1$	Objective social decision: withholding action
VII	(F) (T)	Subjective social decision: action
VIII	$0 < F \mid T < 1$	Objective social decision: action

* Parentheses represent discrete truth values; vertical lines, cuts in the truth or probability continuum.

Analysis and Development

The most general model of our ideas is shown in Table 1, relation I. It states that for any proposition the probability, p, of it being true is some value equal to or greater than zero and equal to or less than one. Classical logic holds, however, that there are two truth values, i.e., false, F, or true, T; as shown in Table 1, relation II. We know now that relation II is relevant to formal propositions only. But relation I implicitly contains relation II since by simple translation we get relation III. That is, the *limits* of the truth or probability continuum suffice for the truth values of classical Aristotelian logic.

But no empirical proposition is ever certain, that is, p can never be 0 or 1 for such propositions because (*a*) there always seems to be error in human observations, (*b*) no empirical proposition is ever regarded as absolutely true or absolutely false by all those competent to judge, i.e., by all scientists at all times, and (*c*) all possible observations never have been and never will be made regarding any empirical matter. Item *c* seems to be true because (*i*) positive or negative instances may occur in the future and (*ii*) even in a specified time interval in which observa-

tions are made, all possible observations are not made nor do all scientists at that time make observations of the specified phenomenon; therefore, negative or positive instances could occur in the present, could have occurred in the past, and could occur in the future. Hence, that part of relation I which is appropriate to empirical propositions is given by relation IV.

In the practical world of affairs, instances constantly arise about which the clinician wishes to postpone judgment or opinion until he has obtained more information, made more observations, consulted with others, and weighed the evidence before he makes a diagnosis, decides on appropriate therapy, or initiates a course of action. When such instances arise, the clinician often makes a subjective analysis of a patient's case history, test results, present behavior, etc. For example, in the case of Dick the clinician's judgment was to postpone or withhold a definite course of action. The empirical proposition on which his decision rests might be something like "Dick is psychotic." At the time a *subjective* social decision had to be made, the proposition "Dick is psychotic" remained in doubt. He requires, therefore, a 3-value logic: false, doubtful, and true. Relation V is the appropriate logical frame of reference in this instance. However, when an acceptable estimate of p can be calculated, three intervals of relation IV can suffice for the three discrete truth values of relation V. The calculated estimate of p would permit an *objective* social decision to be made that definite action should be withheld in Dick's case if p fell within the D-interval of relation VI.

It seems obvious that a 3-value logic is a cautious guide to decision making in a practical world of affairs. It permits us to postpone a decision until we are surer of the factors involved and their relative weights. But sooner or later the clinician is forced by the demands of the society in which he lives and works to make another kind of decision. That is, societal demands often force him to say "yes" or "no"; this will work or that will not work. He must decide whether a person was or was not morally responsible at the time of a crime, whether a patient is or is not a schizophrenic, whether a person is or is not to be committed to a mental hospital, whether a person will or will not violate his parole, whether a patient is or is not to be released from the hospital, etc. The clinician must make many decisions of an "either-or" nature. The statement regarding Harry is of this sort.

At first glance it may seem that a 3-value logic also would be appropriate to a social decision initiating a definite course of action. But there are serious theoretical as well as practical limitations involved. Situations often arise where "close" decisions have to be made. The D-category of relation V or the D-interval of relation VI may be so wide that a proposition must remain forever in doubt. His files would become full of unresolved cases. Before long, social pressures (administrative, parental, legal) would be brought to bear on him to make definite decisions regarding these cases.

The empirical propositions upon which the statement about Harry rests is "Harry is a mental defective." The social rule, say, is that all those classified as mentally defective must be committed to the state institution for mental defectives. The clinician may study at great length all available information regarding Harry and then make the *subjective* social decision that Harry is a mental defective and, therefore, must be committed to the state institution. Relation VII is applicable here. However, when the clinician can calculate an estimate of p, he may define two intervals of relation IV and relation VIII results. Relation VIII is useful in making *objective* social decisions which initiate a definite course of action.

The preceding development indicates that of the many possible frames of reference (1-, 2-, 3-, . . . , n-value), there are three which have particular relevance to clinical statements: 2-value or 2-interval, 3-value or 3-interval, and multivalue. The concept that unifies them all is the truth or probability continuum.

Summary

The clinical psychologist must serve two masters: science and society. His professional statements will fall in three domains: formal, empirical, and societal. Statements with societal significance are often implictily based

upon empirical propositions. Each formal or empirical proposition can be regarded as having a certain probability, p, of being true. The general model is the truth or probability continuum: $0 \leq p \leqq 1$. Clinical statements about a patient are directly or indirectly empirical propositions.

A clinician has two kinds of social decisions to make about a patient: to withhold a course of action or to initiate a definite course of action. His *subjective* social decisions require a 3-value logic (false, doubtful, true), whenever he wishes to withhold action. His *subjective* social decisions require a 2-value logic (false, true) when he must initiate a definite course of action. He can, however, make *objective* social decisions whenever an estimate of p can be calculated. The 3-value and 2-value logics may then be regarded as 3-intervals and 2-intervals respectively of the probability continuum.

Received December 8, 1954.

References

1. Allport, G. W. The psychologist's frame of reference. *Psychol. Bull.*, 1940, **37**, 1–28.
2. Allport, G. W. *The use of personal documents in psychological science.* New York: Soc. Sci. Res. Counc., Bull., 1942, No. 49.
3. Ayer, A. J. *Language, truth and logic.* London: Victor Gollancz, 1949.
4. Snygg, D., & Combs, A. W. *Individual behavior.* New York: Harper, 1949.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Delayed Gratification and Ego Development: Implications for Clinical and Experimental Research

Jerome L. Singer
Franklin D. Roosevelt VA Hospital, Montrose, New York

Since its initial delineation in Freud's *The Interpretation of Dreams* (13), the concept of delayed discharge has played a pivotal role in the psychoanalytic theory of ego development and the psychology of thinking. Little attempt has been made, however, to explore some of the implications of this concept and to deduce testable hypotheses or to survey systematically some of its parameters in relation to views held by other theorists in psychology. The last decade has witnessed a marked increase of interest in ego functions and a consequent coalescence of psychoanalytic theory and general psychology. It now seems possible to attempt at least an initial survey of the scientific usefulness and validity of Freud's view of delay. In the present paper the concept of delay will be related to the views of other theorists, relevant empirical evidence will be cited, and some implications both for further experimentation and practical utility will be suggested.

Origins in Psychoanalytic Theory

The distinction between *primary* and *secondary* processes in mental functioning has been described by Jones (22) as one of Freud's most profound formulations. The shift from the chaotic, associative, diffuse, and "syncretic" character of primary-process thinking to the more synthetic, goal-directed, reality-oriented secondary process involves as a key step the enforced delay or deferment of gratification (13). As Freud described the sequence of thought development, the gradual internalization of the delaying mechanism provided the clue to the emergence of the reality principle (11). The arousal of a need-tension pressing for discharge results in the infant in a restless discharge of motor activity which eventually subsides upon contact with the satisfying object, e.g., bottle or breast. When such restless motor activity has been repeated frequently in association with the satisfying object, the absence in reality of the object will lead to its hallucinatory representation as soon as the motor innervation is initiated by the aroused need. This step is part of the primary process and reflects the operation of the pure pleasure principle. It can be seen clearly in evidence in the autistic perceptions of psychotics, wish-fulfilling dreams, and hallucinatory food responses of starving castaways. Cathectic energies are still mobile and pressing for immediate discharge at this stage; hence, the ease with which things can stand for other things associated with them only by spatial or temporal contiguity. Chaplin's movie, *The Gold Rush*, portrays this most acutely and humorously. The little tramp with his jerky, restless movements is suddenly transformed into a juicy fowl in the eyes of his starving companion in the snowbound cabin.

Freud at no time brought together all the steps involved in the shift from the hallucinatory representation of the absent drive object to the internalized control of discharge through thought process. Nevertheless, the sequence roughly gleaned from a number of different papers seems to involve a gradual "neutralization" of libidinal energy through identification and incorporation of parental figures, transformation of mobile energies

into "bound cathexes," and formation of a functional substructure of the personality, the *ego*, whose properties include motor control, purposive action, directed thought, reality-cognition, and various defensive measures which organize or segmentalize the diffuse innervations arising from basic drives (id) (9, 10, 11, 12, 13).[1] As early as 1900 Freud had already suggested that the transition from mobile to bound cathexes made possible the use of small, controlled energy expenditures at the behest of the secondary process which could serve as signals (13), or as he put it later, "experimental actions" (11). In effect the secondary process substitutes a "thought identity" for the hallucinatory wish fulfillment. Freud wrote, "All thinking is no more than a circuitous path from the memory of a satisfaction (a memory which has been adopted as a purposive idea) to an identical cathexis of the same memory which it is hoped to attain once more through an intermediate stage of motor experiences" (13, p. 602).

Subsequent psychoanalytic theorists have not altered Freud's concepts to any great degree. A brief but stimulating paper by Ferenczi (7) carries Freud's concepts to the clinical sphere and points to some individual differences in motor activity which are coordinated with differences in thought production. Ferenczi cites an example of a patient whose constant foot wiggling would stop whenever she began to think consciously. It became possible for the analyst to use this cessation of motor activity as an indication of resistance if the patient failed to communicate verbally the thoughts that had followed upon her decreased motor response. Fenichel (6), too, has elaborated on Freud's conceptions without, however, altering their sequence or structure to any significant degree, while Anna Freud (8) in her work has emphasized what Hartmann (16) would call the "conflictual ego-sphere" and deals with fantasy (denial) or hypercathected thought processes (intellectualization) only in their defensive role.

The recent work of Hartmann (16), Federn (5), Weiss (56), Hartmann, Kris, and Lowenstein (17), Kris (24), and Rapaport (39, 40) has added new breadth to Freud's conception of the ego. Rapaport (39), in particular, has taken an extremely significant step in the direction of providing a systematic, scientifically verifiable formulation of Freud's concept of delayed discharge. He traces the meaning of this concept at different vantage points: *genetic* (primary to secondary process), *topographic* (unconscious or preconscious and conscious), *structural* (ego functions), *economic* (mobile to bound cathexes or delibidinized energies), and *biological* (adaptive nature of delay and thought). By this analysis it becomes possible to see more clearly the relationship of the psychoanalytic theory of thinking to the work of others less oriented toward concern with the "neutralization" of instinctual energies but more with the adaptive function of thought in biological survival and interpersonal relationships. Rapaport writes, "Thinking explores the possible pathways of action to find the one of least resistance, least danger, and greatest directness, while preserving almost intact the energy necessary for motor action" (39, p. 169). This formulation, in essence unchanged from Freud's formulations in 1900 and 1911, is nevertheless an advance in clarifying the concept of delay without the confusing terminology rooted in Freud's vaguely defined cathectic dynamics. Rapaport's huge volume (40) contains a challenging presentation of the hierarchical, almost functionally autonomous structures of motives and thought processes in the personality structure.

Other Viewpoints of Delay and Thought

The belief that conscious thought or fantasy bears some relation to the inhibition of an overt motor response directed toward some sort of gratification has appeared in the writings of many investigators who worked more or less independently of psychoanalytic theory. The early American functionalists and behaviorists held similar views, even if we except the somewhat extreme and experimen-

[1] The meaning of the concept of free and bound cathexes, attributed by Freud to Breuer, is never completely clear in Freud's writings. Breuer (3) seems to have linked "bound cathexes" to "tonic excitation of nervous tension," a phrase of interest chiefly because of its at least superficial similarity to Werner's "sensory-tonic theory of perception" (58).

tally unsupported approach that thought was *nothing more* than implicit movement of the speech mechanism. James (21) saw "verbal imagination" as the outcome of "suppressed articulation" while Dewey (4) claimed that thinking becomes intensified upon the interruption of a usual sequence of motor acts. Washburn (55) held a somewhat similar point of view, noting also that consciousness and "associative activity" diminish with the overt performance of the act. E. B. Holt, an early pioneer in the integration of Freudian theory with the emerging American behaviorism, wrote in 1915:

> Thought is . . . more than the thing thought about: there is active thinking about objects. If we look once more at the least manifold in which cognition occurs—a living organism in, and responsive to, an environment—we see that this further element is the active play of motor attitude, which eventually resolves itself into the less labile but more forceful phenomenon, conduct. Thus, thought is latent course of action with regard to the environment . . . , or a procession of such attitudes (19, page 98).

The theories of the behaviorist and functional psychologists, while less comprehensive in their implications than the psychoanalytic theory, have led at least to some experimental research. These investigations, as summarized by Vinacke (54), have indicated that thought is indeed accompanied by some implicit muscular activity albeit by no means has it been shown to be clearly the *result* of such activity.

More directly related to the psychoanalytic theory, both in influence and implications for a personality theory, are views held by important workers in developmental psychology like Piaget (36), Werner (60), Lewin (29), Murphy (35), and Luria (30). All of these investigators present empirical data to support a view that thought involves the gradual interiorization of motor response or speech which must be inhibited from overt expression because of social restraints or pressures. Luria (30), for example, describes the change with increasing age of children from the vigorous acting out of a battle scene to its representation in drawing with minimal accompanying speech or motor activity. Piaget (36) has traced the change in children's language from the egocentric comments in which everything that comes to mind is pronounced aloud without regard for its implications. Socialization enforces inhibition of a great deal of this speech pattern with the consequent heightening of thought activity, conscious and unconscious. Wolff and Precker (64) describe the sequence of development in terms of the changing reaction of the growing child in response to being struck: at one stage he responds with a direct counterblow, later he may verbally abuse the assailant, and finally he may completely inhibit or delay overt response but fantasize various countermeasures.

Perhaps most precise in its derivation and fruitful in terms of experimental research relevant to the Freudian concept has been the work of Werner (58) and Werner and Wapner (61, 62, 63). Building in part on earlier physiological investigations as well as upon Werner's (60) extensive survey of developmental patterns, these writers have formulated a sensory-tonic theory of perception which offers at least the possibility of a more precise ultimate formulation of Freud's view of delayed discharge in operational, experimentally verifiable terms. Werner and Wapner describe perception as a function of the dynamic interaction of externally derived sensations and pre-existing tonic states in the organism. Both sensory and tonic events have equivalent functions in the economy of the organism and may serve vicariously as substitutes for each other. When available energy is not released in bodily movement the tonicity will express itself in perceptual displacement, e.g., illusory motion, or, perhaps one might add, "movements in the mind's eye," fantasies of motion. There has been an increasing body of experimental literature in support of such a theory of the "vicarious channelization" (62, 63). As an example, the finding reported by Goldman (15) may be cited. Subjects responded to the autokinetic phenomenon under *immobilization, control,* and *heightened motor activity* conditions. Readiness to perceive motion was greatest under immobilization, less under a control condition, and least during activity. The theory and experiments of Werner and his colleagues thus offer a challenging opportunity, but their demonstration of blocked motor

discharge leading to perception of movement must still be related more clearly to heightened imaginativeness if the Freudian concepts are to be assimilated into their structure.

Experimental Research with Rorschach's Movement Responses

Perhaps the most striking possibility for operational expression and experimental study of the concept of delay has come in research with the Human Movement response of the Rorschach test. Rorschach (41) in his studies of responses to inkblots delineated a type of response which involved perception of humans in action (*M*). He dealt at great length with his observations that persons who saw several such *M* responses on a blot series tended, on the one hand, to be relatively stable or inhibited in motor activity and, on the other hand, rather imaginative, given to considerable inner living and creativeness or planfulness. In effect, Rorschach, although lacking a clear-cut theory, was formulating a triadic linkage of motor inhibition, perception of motion, and imagination or inner life. Here then appears a possible bridge between the psychoanalytic concept of delayed gratification (inhibition of motor response) and the sensory-tonic theory of vicarious functioning of motor activity and motion perception. The relationship has indeed been emphasized by Rapaport's (40) concern with the delaying capacity of *M*-type individuals and by Werner's (59) work comparing Rorschachs of hyperactive and motorically-phlegmatic feebleminded children.

The heuristic value of Rorschach's *M* response lies in two separate if ultimately related directions. On the one hand we may attempt to test both Werner's and Freud's hypotheses fairly directly by observing what effects emerge when motion is inhibited prior to or during a Rorschach examination. If this delayed discharge of motor activity has the implications described above then, an increase in *M* responses should be a consequence of this inhibition. On the other hand, on the assumption of individual differences in ego organization and control, it is likely that long-standing patterns of motor inhibition or hyperactivity will be reflected in differential Rorschach patterns. Thus, studies of subjects who contrast sharply in Rorschach *M* or in imaginativeness or motor control may throw valuable light on the nature of the delaying mechanism. The first experimental approach is chiefly of theoretical interest and ultimately may lead to the crucial experimentation upon whose results the theory of delay and thought may stand or fall. The second hypothesis, while less crucial theoretically, has important practical implications including the possible development of measuring devices for the vital ego function of delay of gratification and effective use of planning and imagination.

Turning first to the direct consequences of motor inhibition or delay, two experimental studies have reported increases in Rorschach movement responses of normals following periods of motor inhibition, i.e., standing rigidly in awkward postures or voluntary slow writing (32, 48). Another study has investigated the effects of speeding up or slowing down tempos of Rorschach response and has yielded the conclusion that production of *M* responses is indeed related to tendencies or capacities to delay response (45). Further work remains to be done but it seems clear that so far the linkage of delay of motor discharge and motion perception and imagination can be demonstrated directly in normal adults. No work with children has yet been carried out in this direction. A number of studies do suggest that the *M* response emerges most clearly in age groups that are being subjected to the first striking enforced socialization and motor restraint of school entrance (28, 58). It can be said that Freud's early formulation of thought as arising from delayed gratification and discharge of minute quantities of energy in experimental action is certainly consistent with these findings. Whether the subtler refinements of the theory such as the transformation from mobile to bound cathexes or the desexualization of energy leading to hypercathecting of ideational representations can be expressed in operational terms and tested experimentally remains to be seen. Holt (20) has suggested some possibilities in this connection.

The second general avenue of exploration in this field has been more thoroughly traversed because it calls for correlational pro-

cedures rather than direct experimental manipulation. Can it be shown that individuals who are given to or capable of extensive impulse control, deferment of gratification or motor control (apart, that is, from athletic skill), will reflect correspondingly both low thresholds for perception of motion and greater imaginativeness? In effect this implies that some people do manifest a persistent style of coming to terms with their environment which involves relative delay of immediate reactions to needs or provocative stimulation and employment of planfulness or fantasy in their mode of relating. Rorschach's distinction between the introversive and extratensive experience-types is clearly an example of such a cognitive style and the testimony of various experienced Rorschach authorities certainly confirms this dimension of personality variation (2, 23, 38, 42).

Actual systematic investigations have linked characteristic modes of motor reactivity to Rorschach movement responses (1, 59). It has also been observed that both normal and schizophrenic subjects who yield numerous *M* responses are more capable of inhibiting motor responses on slow-writing tasks or are less active and responsive to environmental stimuli while observed in a waiting room (32, 47, 49). The other aspect of this relationship, *M* and creativity or imaginativeness, cannot be studied directly. Measures such as the ability to transcend objective stimulus material in thematic apperception stories which may be considered reasonable approximations of this fantasy capacity, have, however, uniformly revealed significant correlations between *M* and imaginativeness, even with general intelligence controlled (1, 18, 37, 43, 44, 47, 53). Recently Singer and Herman (47) reported data revealing that schizophrenics differing in *M* responses differed correspondingly in a series of situations tapping motor control, imaginativeness, spontaneous motor activity, and impulsivity. The results suggested that a common factor does indeed link motility, motion perception, and thought as characteristic ego functions. Motor inhibition has also been related to cognitive inhibition, the tendency to delay verbalization of an old association and to rapidly substitute a new one (31). Finally, some recent work has shown Rorschach's *M* to be related to high levels of aspiration, planning ability, and impulse control (51).

It should be noted, however, that none of the linkages reported in this section represent extremely high correlations. As a matter of fact, considerable variability is present and it is certain that numerous factors ranging from influences of other spheres of ego functioning (16) and from direct situational factors in the face-to-face subject-examiner relationship modify the emergence of the cognitive style in question. It may be concluded, nevertheless, that in general the capacity or tendency to delay response or inhibit motor discharge appears to be a characteristic adjustment of many persons and it is linked in them with tendencies to be sensitive to movements on ambiguous forms or to engage in fantasy activities with relative ease.

Some Genetic Considerations and Clinical Implications

Accepting at least temporarily the last conclusion, what can be said of the way in which these characteristic modes of coming to terms with the environment originate? It seems likely that many temperamental characteristics, particularly gross features such as relative motility tendencies (in which large differences emerge in children [57]) may be constitutionally based. Certainly Freud (12) came to feel that quantitative factors in libido distribution were of such a nature. Nevertheless, one is tempted to seek for considerable additional sources of individual variability in acquired characteristics, modes of adjustment, defensive techniques, or values learned from significant interpersonal situations in the family constellation. One may well ask, what sort of relationships in the family setting foster in the child a capacity to defer gratification or to temporarily restrain motor discharge in response to a need with subsequent resort to the realm of imagination or trial action? A clue is forthcoming in Freud's stress on identification as a means of neutralizing libido and establishing ego defenses and countercathexes. In family constellations where relatively strict patterns of eating, toileting, washing, napping, and recreation are the rule children may more readily, in order to master

their tensions, learn to accept delay since the sequence or pattern is consistent. In thus delaying and striving to imitate the adults they find that resort to the level of fantasy gratification is possible and is frequently rewarded by parental figures or leads to new possibilities for mastering the reality situation. Thought and delay become valued (cathected) and organized and a new hierarchical structure of personality emerges, Lewin's level of irreality (29). Within reasonable limits this pattern is highly adaptive and, indeed, one of the most uniquely human characteristics for it frees man from dependence upon immediate stimulation. Yet it can be carried to excess, as in the case where the extreme anxiety of parents leads to overly rigid patterns with the result that unceasing resort to inhibition and delay leads the child into obsessional rumination that gradually loses its reality orientation and creativity.

Lair (26) has presented a review of considerable theoretical material and some data which support the view that an opportunity for identification with benign and loving adults in very early childhood is essential to the development of capacities for motor inhibition and planful fantasy. Goldfarb's studies of adolescents whose first three years of life had been spent in institutions where close relationships with adult figures were not possible are highly suggestive in this connection. These institutional children showed almost uniformly a thinking disorder characterized by lack of time perspective, hyperactivity, absence of "normal inhibitory pattern," and lack of planfulness (14). Kupper (25) has traced in detail how certain types of family constellation lead to development of strong intellectual values and inhibited motor patterns. The work of Mowrer (34) on language development has represented perhaps the first systematic attempt to develop more precisely the manner in which symbolic behavior is learned through attempted identification with the secondary reinforcing characteristics of benign parental figures. Mowrer's formulations in the framework of two-factor learning theory certainly afford the possibility of expressing some of the analytic concepts of identification and "hallucinatory representation" in operational terms and may open this area to experimental study.

To some extent, it would appear, then, that the varieties and opportunities for parental identification may play a role in the capacity to develop delaying mechanisms and suitable imaginative processes. To the extent that social class factors play a role in the organization of family constellations these factors may markedly influence the relative resort to deferred gratification patterns. The middle class family with its routines, family unity at mealtimes, and general emphasis on saving or education provides the child with a setting that fosters such delay. This is particularly the case since the consistency of the pattern leads to the relative assurance that satisfaction will come eventually. In the lower class family parents are less available (particularly fathers), meals are haphazard with no assurance of their being forthcoming, and the value of delay is less clear to the child. Hence we may anticipate that lower class children will show different patterns of means-end cognizance, time evaluation, and imaginativeness. Some tendencies of this sort have indeed been reported in the recent literature (27, 35, 46).

Returning to the Rorschach movement response as a possible measure of delaying capacity and fantasy we might expect that high-*M* subjects would show more positive feelings toward parents or at least greater identification with them. Studies with the TAT have offered some supporting results but considerably more work remains to be done here (44, 50). It would be important to study known family constellations and observe the patterns of delay, motor inhibition, or fantasy which emerge in the children. One possibility might be exploration of cultural groups known to differ markedly in family structures, e.g., Irish and Italian lower class.

In conclusion, then, it seems plain that Freud's original concept of delayed gratification leading to thought as experimental action is indeed a fruitful one. It is supported in rough outline at least by the observations of a host of investigators and by considerable experimental research particularly that with the Rorschach *M* response. The dimension of motor inhibition—motion perception—fantasy

seems a strikingly useful one, capable of empirical study and providing an interesting approach to the study of personality structure in the family setting. The concept of delay or impulse control as a basic ego function affords also a highly significant contact point between psychoanalytic theory and general psychology. More than fifty years after its original pronouncement, Freud's theory of thought development is vigorously alive and beckons toward new vistas of research.

Summary

This paper represents an initial attempt to place Freud's concept of delayed gratification and thought development in the perspective of a broader framework of theory and research in psychology. After tracing the development of this concept in psychoanalytic literature, points of contact with early American behaviorist and functional psychological points of view were noted. Experimental research that bore on the theory was cited with particular emphasis on Werner's sensory-tonic theory of perception and the empirical data which has grown out of Rorschach's linkage of motor inhibition, motion perception, and imagination. Some implications of this concept for the study of identification processes, internal control mechanisms, and differential parental influences in the family constellation were suggested.

Received November 5, 1954.

References

1. Barron, F. Threshold for perception of human movement in inkblots and its personality correlates. *Amer. Psychologist*, 1953, **8**, 317–318. (Abstract)
2. Beck, S. J. *Rorschach's test*. Vol. II. New York: Grune & Stratton, 1946.
3. Breuer, J., & Freud, S. Studies in hysteria. In *The standard edition of the complete psychological works of Freud*. Vol. 2. London: Hogarth Press, 1952.
4. Dewey, J. *How we think*. New York: Heath, 1933.
5. Federn, P. *Ego psychology and the psychoses*. New York: Basic Books, 1952.
6. Fenichel, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.
7. Ferenczi, S. Thinking and muscle innervation. In *Further contributions to the theory and technique of psychoanalysis*. New York: Basic Books, 1952. Pp. 230–232.
8. Freud, Anna. *The ego and the mechanisms of defense*. New York: International Universities Press, 1946.
9. Freud, S. *The ego and the id*. London: Hogarth Press, 1923.
10. Freud, S. *The problem of anxiety*. New York: Norton, 1936.
11. Freud S. Formulations regarding the two principles in mental functioning. In *Collected papers*. Vol. IV. London: Hogarth Press, 1946. Pp. 13–21.
12. Freud, S. *An outline of psychoanalysis*. New York: Norton, 1949.
13. Freud, S. The interpretation of dreams. In *The standard edition of the complete psychological works*. Vol. V. London: Hogarth Press, 1953. Pp. 339–628.
14. Goldfarb, W. Psychological privation in infancy and subsequent adjustment. *Amer. J. Orthopsychiat.*, 1945, **15**, 249–254.
15. Goldman, A. Studies in vicariousness: degree of motor activity and the autokinetic phenomenon. *Amer. J. Psychol.*, 1953, **66**, 613–617.
16. Hartmann, H. Comments on the psychoanalytic theory of the ego. In *The psychoanalytic study of the child*. Vol. V. New York: International Universities Press, 1950. Pp. 74–96.
17. Hartmann, H., Kris, E., & Lowenstein, R. Comment on the formation of psychic structure. In *The psychoanalytic study of the child*. Vol. II. New York: International Universities Press, 1947. Pp. 11–38.
18. Hays, W., Gellerman, S., & Sloan, W. A study of the relation between the verb-adjective quotient and the Rorschach experience-balance. *J. clin. Psychol.*, 1951, **7**, 224–227.
19. Holt, E. B. *The Freudian wish and its place in ethics*. New York: Holt, 1915.
20. Holt, R. Implications of some contemporary personality theories for Rorschach rationale. In B. Klopfer, Mary D. Ainsworth, W. G. Klopfer, & R. Holt (Eds.), *Developments in the Rorschach technique*. Vol. I. Yonkers, N. Y.: World Book, 1954.
21. James, W. *The principles of psychology*. New York: Holt, 1890.
22. Jones, E. *Life and work of Sigmund Freud*. Vol. I. New York: Basic Books, 1953.
23. Klopfer, B., & Kelley, D. M. *The Rorschach technique*. Yonkers, N. Y.: World Book, 1942.
24. Kris, E. On preconscious mental processes. *Psychoanal. Quart.*, 1950, **19**, 540–560.
25. Kupper, H. L. Psychodynamics of the "intellectual." *Int. J. Psychoanal.*, 1950, **31**, 85–94.
26. Lair, W. S. The psychoanalytic theory of identification. Unpublished doctor's dissertation, Harvard Univer., 1949.
27. Levi, J., & Kraemer, Doris. Significance of a preponderance of human movement responses in children below age ten. *J. proj. Tech.*, 1952, **16**, 361–365.
29. Lewin, K. *A dynamic theory of personality*. New York: McGraw-Hill, 1935.

30. Luria, A. P. *The nature of human conflicts.* New York: Liveright, 1932.
31. Meltzoff, J., & Levine, M. Motor and cognitive inhibition. *J. consult. Psychol.,* 1954, **18,** 355–358.
32. Meltzoff, J., Singer, J. L., & Korchin, S. J. Motor inhibition and Rorschach movement responses: a test of the sensory-tonic theory. *J. Pers.,* 1953, **21,** 400–410.
33. Mitchell, H. E. Social class and race as factors affecting the role of the family in Thematic Apperception Test stories of males. Unpublished doctor's dissertation, Univer. of Pennsylvania, 1950.
34. Mowrer, O. H. *Learning theory and personality dynamics.* New York: Ronald Press, 1950.
35. Murphy, G. *Personality.* New York: Harper, 1947.
36. Piaget, J. *The language and thought of the child.* New York: Harcourt, Brace, 1932.
37. Pickering, W. D. A comparison of predominant verbal levels on the Thematic Apperception Test with Rorschach experience-balance. Unpublished doctor's dissertation, Univer. of Pittsburgh, 1950.
38. Piotrowski, Z. A. A Rorschach compendium, revised and enlarged. *Psychiat. Quart.,* 1950, **24,** 543–596.
39. Rapaport, D. On the psychoanalytic theory of thinking. *Int. J. Psychoanal.,* 1950, **31,** 161–170.
40. Rapaport, D. *Organization and pathology of thought.* New York: Columbia Univer. Press, 1951.
41. Rorschach, H. *Psychodiagnostics.* New York: Grune & Stratton, 1942.
42. Schachtel, E. G. Projection and its relation to character attitudes and creativity in the kinesthetic responses. *Psychiatry,* 1950, **13,** 69–100.
42. Schumer, Florence. Some behavioral correlates of Rorschach human movement responses. Unpublished doctor's dissertation, Yale Univer., 1949.
44. Shatin, L. Rorschach adjustment and the Thematic Apperception Test. *J. proj. Tech.,* 1953, **17,** 92–101.
45. Siipola, Elsa, & Taylor, Vivian. Reactions to inkblots under free and pressure conditions. *J. Pers.,* 1952, **21,** 22–47.
46. Singer, J. L. Projected familial attitudes as a function of socio-economic status and psychopathology. *J. consult. Psychol.,* 1954, **18,** 99–104.
47. Singer, J. L., & Herman, J. Motor and fantasy correlates of Rorschach human movement responses. *J. consult. Psychol.,* 1954, **18,** 325–331.
48. Singer, J. L., Meltzoff, J., & Goldman, G. D. Rorschach movement responses following motor inhibition and hyperactivity. *J. consult. Psychol.,* 1952, **16,** 359–364.
49. Singer, J. L., & Spohn, H. E. Some behavioral correlates of Rorschach's experience-type. *J. consult. Psychol.,* 1954, **18,** 1–9.
50. Singer, J. L., & Sugarman, D. Some Thematic Apperception Test correlates of Rorschach human movement responses. *J. consult. Psychol.,* in press.
51. Singer, J. L., & Wilensky, H. Delaying capacity, fantasy, and planfulness. Paper read at East. Psychol. Ass., Philadelphia, 1955.
52. Thetford, W. N. Fantasy perceptions in the personality development of normal and deviant children. *Amer. J. Orthopsychiat.,* 1952, **22,** 542–550.
53. Vernier, Claire, M., & Kendig, Isabelle V. Analysis of the relationships between various measures of creative productivity in two projective tests. *Amer. Psychologist,* 1951, **8,** 349. (Abstract)
54. Vinacke, W. E. *The psychology of thinking.* New York: McGraw-Hill, 1952.
55. Washburn, M. A system of motor psychology. In C. Murchison (Ed.), *Psychologies of 1930.* Worcester, Mass.: Clark Univer. Press, 1930. Pp. 81–94.
56. Weiss, E. *Principles of psychodynamics.* New York: Grune & Stratton, 1950.
57. Wenger, M. A. An attempt to appraise individual differences in level of muscular tension. *J. exp. Psychol.,* 1950, **29,** 379–390.
58. Werner, H. Motion and motion perception: a study on vicarious functioning. *J. Psychol.,* 1945, **19,** 317–327.
59. Werner, H. Perceptual behavior of brain-injured mentally deficient children. *Genet. Psychol. Monogr.,* 1945, **31,** 51–110.
60. Werner, H. *The comparative psychology of mental development.* Chicago: Follett, 1948.
61. Werner, H., & Wapner, S. Sensory-tonic field theory of perception. *J. Pers.,* 1949, **18,** 88–107.
62. Werner, H., & Wapner, S. Toward a general theory of perception. *Psychol. Rev.,* 1952, **59,** 324–333.
63. Werner, H., & Wapner, S. Organismic theory and perceptual response. Paper read at East. Psychol. Ass., Boston, April, 1953.
64. Wolff, W., & Precker, J. A. Expressive movement and the methods of experimental depth psychology. In H. H. Anderson & G. L. Anderson (Eds.), *An introduction to projective techniques.* New York: Prentice-Hall, 1951. Pp. 457–497.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Testing Behavior Hypotheses with the Rorschach: An Exploration in Validation[1]

Arthur E. Eschenbach
Officer Education Research Laboratory, U. S. Air Force

and Edgar F. Borgatta
Russell Sage Foundation

Interest in the Rorschach as reflected in the literature, both in theoretical formulation and report of empirical findings, is extensive. Systematic empirical approaches, however, have been few, and are primarily those involving factor analysis of Rorschach scores (6, 11), or of these scores in conjunction with other scores such as intelligence test performance (10). The scoring categories or diagnostic signs which have been included in the matrix of intercorrelation have not been identical, and where there has been replication of the study, at least reasonable replication of factor structure has been found (10). Factors identified thus far have tentatively been identified with such names as productivity, movement, lack of perceptual control, shading, intelligence, organic pathology, loose conceptual thinking, color awareness and color use, popular response, psychoneurosis, and confused succession. What factors are found is determined in the limit of the approach by the items which are included in the matrix, and the subsections of a population which are included.

Wittenborn (12) has indicated the relevance of a validation study with discrete Rorschach scoring variables, and Wittenborn and Holzberg (13) have presented the complementary argument for using discrete descriptive or behavioral criteria of validation. Essentially, the point is that unless some common variance is found between discrete scoring categories and discrete descriptive or behavioral criteria, there can be little optimism for the diagnostic use of the scoring categories involved. Wittenborn and Holzberg (13) found rather discouraging results in relating Rorschach scoring categories to symptom manifestations in descriptive diagnosis.

Validation of a diagnostic tool should be in terms of the kinds of uses to which it is put. In this study we are interested only in certain aspects of scoring which are defined in terms of the responses of the subject rather than the general impression or evaluation of the interviewer. The use of the Rorschach as a medium of clinical interview is to be distinguished from its use as a diagnostic tool in itself. That is, the Rorschach is a medium of clinical interview when it is used to establish rapport to observe the behavior of the subject or to evoke otherwise pertinent information which might be awkward to elicit in a formal interview. We are not interested in the characteristics which are most closely associated with the sensitivity and intuitive ability of the

[1] This research was supported in part by the United States Air Force under contract No. AF 33 (038)–12782 monitored by the Human Resources Research Institute, Maxwell Air Force Base, Alabama. Permission is granted for reproduction, translation, publication and disposal in whole or in part by or for the United States Government. Personal views or opinions expressed or implied are not to be construed as necessarily carrying the official sanction of the Department of the Air Force. The research was conducted through the Laboratory of Social Relations, Harvard University. A portion of the data utilized was derived from a dissertation submitted by the senior author in partial fulfillment of the requirements for the Ph.D. degree at the University of Florida. We are grateful to the HRRI staff which assisted in various stages of this research.

interviewer, but with the scores which are most arbitrarily defined and reproducible.

Procedure

Studies which measure the difference of means of selected Rorschach scores among extreme subsections of the population may prove significant. The information may, however, be of little value for prediction. That is, since two special cells are compared, or one special cell is compared to the remainder of the population, it may be that estimates of the amount of variance actually under control will indicate that the relationship is too weak to be practically useful. In this study we are interested in measuring the relationship of Rorschach scores and behavioral descriptions which exist in a relatively broad population.

Subjects. Our sample consisted of 125 ostensibly normal male Air Force enlisted personnel representing the seven enlisted grades. The subjects were normally distributed on the SRA Test of Primary Mental Abilities, had an average age of 27 years with a range from 20 to 45 years, and were also heterogeneous in regard to education and military occupation. The only apparent nonnormal characteristic of this population was that a majority of the 18 Basic Airmen had recently been incarcerated for breaches in discipline and had lost whatever military grade they had.

Rorschach scoring categories. Since the frequency in some of the commonly used Rorschach scoring categories is extremely small, and the scores are discrete and unitary, these scoring categories cannot properly be utilized in testing relationships among persons. That is, the units of scoring in these categories are too large for the population to be distributed according to ordinary correlational requirements. Split-half reliability of this type of item will almost invariably be negative as a function of the distribution of the scores. For this reason, we have taken scores by the frequency of occurrence without regard to smallest detail specification, e.g., we take *C* to mean the sum of all cases in which color occurs, such as *C, CF, FC, FCV,* etc. Our scoring categories in this study are: number of responses (*R*); $F+/(F+)+(F-)$; total color responses (*C*); total motion (*M*); total vista (*V*); total shading plus texture ($T+V$); total space (*S*); total large detail (*D*); total rare detail (*Dd*); popular responses (*P*); organization (*Z*); *content variety;* animal plus animal detail ($A+Ad$); anatomy (*An*), total whole (*W*).

Criteria. We have taken the fifteen scoring categories outlined above and have explored their relationship with five groups of variables selected as valid criteria: (*a*) scores of performance in interaction; (*b*) ratings by superiors; (*c*) primary mental ability scores; (*d*) sociometric status indices; and (*e*), background information.

The scores of performance in interaction were taken in a "small group" experiment which is reported elsewhere (4). The scores represent the average number of acts of the person in four sessions which involve getting acquainted, planning role-playing sessions, and free discussion. In each session, the person participated with two different persons. The interaction observed in each session was for a period of 24 minutes, and represented parts of the session in which the subjects were presumably unaware of being observed. The categories of scoring used were those of interaction process analysis (1), which have been used extensively for behavioral description of group and individual performance in discussion and problem-solving groups. Here we include the *initiated behavior* in the twelve basic scoring categories, four subdivisions of the twelve categories, and the total.

Briefly, the *categories of this observation system* may be identified as follows: categories 1, 2, and 3, showing solidarity, tension release, and agreement; categories 4, 5, and 6, giving suggestion, opinion, and orientation; categories 7, 8, 9, asking for orientation, opinion, and suggestion; and categories 10, 11, and 12, showing disagreement, tension, and antagonism. The first three and last three categories are known as the positive and negative social emotional categories, respectively. The central six categories are known as the task categories.

The *ratings by superiors* were made by the highest ranking commissioned officers having direct or effective contact with the airmen. The ratings were in terms of performance variables judged particularly important in the

military context: ability to understand the situation and take necessary initiative; ability to handle men and make himself understood by his men; knowledge of the technical aspects of his job; personal conduct and bearing; and general over-all rating.

The scores of *primary abilities* included are those from the SRA Primary Mental Abilities: verbal meaning, space, reasoning, number, word fluency, and a total quotient.

The *affective sociometric indices* are: propensity to be a chooser, propensity to expect to be chosen, choices received (popularity), expectancy of others that one will be a chooser. These were based on a criterion of enjoyment of participation in the session completed. Two other indices on a criterion of amount done to determine the way the session went are also included: chosen as a leader, and chooses self as a leader.

The *background information* includes age, educational level, and the subject's military grade.

Method of Analysis

The method of analysis was governed by experience in the search of empirical hypotheses concerning Rorschach variables.[2] The first question concerned the relationship of R, the number of responses, to the criteria. The second question concerned the relationship of the other Rorschach variables to the criteria. However, since we were primarily interested in the relationships of the other variables not attributable to the number of responses, our matrix was one of partial correlations (Table 1). We were further interested in whether the correlations tend to be ordered properly in regard to behavioral hypotheses which are phrased according to the criteria included. Even if the degree of relationship involved is low, the proper ordering of relationships according to hypotheses would indicate the feasibility of a multiple regression (configurational) approach. Thus we stated forty-four hypotheses. The hypotheses were then tested on the basis of the expected coefficient signs of the relationships of the Rorschach scoring categories with the multiple criteria.

Our definitions of variables for the tests of hypotheses were established before examination of the data.[3] We do not have space for justification of the *behavioral* criteria, which do not have obvious face validity in all cases.

Testing the Hypotheses

In the tests of hypotheses, for economy in presentation, we provide the information in Table 2 in the following order: (*a*) The hypothesis is stated. (*b*) At least one source of the statement is indicated. (*c*) The identification of the variable(s) in Table 1 to which the Rorschach variable is to be compared is stated. (*d*) The *expected* sign of correlation under hypothesis is indicated. (*e*) The *actual* sign of the correlation in Table 1 is given. (*f*) Hypotheses are judged as: rejected (significantly in the wrong direction); in the wrong direction; in the right direction; not rejected (significantly in the right direction). Where two or more possible tests are available, they are given consecutively. For the two-tailed test with .05 level, absolute $r = .174$ is significant (italicized in Table 1), which is equivalent to the .025 in hypothesis testing.

Summary and Interpretation

Of five hypotheses examined concerning the meaning of the total number of Rorschach responses, one was definitely not rejected and the other four were ordered in the right direction.

Of thirty-nine hypotheses other than those concerning the total number of Rorschach responses, with forty-four possible tests, there were: two nonrejection, twenty-one ordered in the right direction, twenty ordered in the wrong direction, and one rejection.

If hypotheses concerning *intelligence only* are considered, there are (including total R): three nonrejections, four in the right direction, and two in the wrong direction.

While it is interesting to consider statements concerning the need for "configurational" in-

[2] Generalizations of relationship of Rorschach scoring categories to personality characteristics have been compiled from the five sources (2, 3, 7, 8, 9) listed in the references. These generalizations have been published separately (5). The generalizations taken for hypothesis test here are those considered most clearly stated in relatively systematic texts.

[3] We are extremely grateful to Robert F. Bales for his help in choosing the interaction behavioral criteria.

Table 1

Intercorrelation of Rorschach Variables with Criterion Variables, Partialled Through *R*

Item	*R*	$\frac{F+}{(F+)+(F-)}$	*C*	*M*	*V*	*T*+*Y*	*S*	*W*	*D*	*Dd*	*P*	*Z*	Content variety	*A*+*Ad*	*AN*
Mean of Rx variables	24.4	85.4	2.6	11.1	0.8	2.1	2.4	5.3	15.1	1.6	5.3	23.3	9.2	10.6	1.6
SD of Rx variables	14.4	9.9	2.7	2.6	1.3	2.3	2.8	3.6	10.1	3.7	2.2	17.0	3.9	6.1	1.7
Number of Rorschach responses	1.000	−.085	*.391*	*.443*	*.230*	*.555*	*.547*	*.188*	*.935*	*.723*	*.515*	*.332*	*.805*	*.828*	*.365*
Ratings by superiors:															
a. Ability to understand	.116	−.040	−.003	*.188*	.014	*.219*	.045	.060	−.069	−.036	.127	.170	−.058	−.049	.029
b. Ability to handle men	.061	.014	.071	.128	.049	.094	−.027	.096	−.045	−.041	.170	*.176*	−.100	.020	.151
c. Knowledge of job	.090	−.034	.028	.089	.100	.165	.118	.062	−.094	−.068	.004	.169	−.018	.009	.051
d. Personal bearing	.084	−.068	−.009	.097	.088	.064	−.062	.031	.004	.008	.148	.094	−.070	−.007	*.178*
e. Over-all rating	.111	−.011	.030	*.174*	.063	.120	.020	.071	−.070	−.025	.100	.167	−.065	−.020	.086
S R A primary mental ability:															
f. Verbal meaning	*.174*	.015	*.202*	.166	−.013	.023	.154	.145	−.094	*−.213*	.154	*.275*	.062	−.009	−.027
g. Space	.155	.026	.118	−.053	−.010	.054	.057	.045	−.011	−.095	.008	.088	−.122	.092	−.046
h. Reasoning	.136	.063	.150	.122	.094	.035	*.197*	.172	*−.183*	−.160	.048	*.251*	−.015	.001	−.091
i. Number	*.182*	.041	−.051	.130	.005	.037	−.048	−.084	.040	−.037	−.009	.078	−.152	.054	−.005
j. Word fluency	*.245*	*.193*	.171	.073	.032	.012	.048	*.187*	−.122	−.133	*.200*	*.229*	*.175*	−.063	−.034
k. Total quotient	*.220*	.051	.106	.144	−.006	.023	.068	.142	−.086	−.143	.108	*.214*	−.088	.079	−.087
Sociometric ratings:															
l. Chosen as a leader	.127	−.050	.171	.014	.092	*.221*	.115	*.228*	−.170	*−.187*	−.093	*.223*	.020	−.145	.136
m. Chooses self as a leader	.156	−.054	.128	.047	.071	.036	.069	.131	−.131	−.055	.009	*.178*	*.188*	*−.221*	−.030
n. Choices made (expansiveness)	−.087	.037	−.166	.081	−.118	.004	*−.251*	−.107	.123	*.199*	−.170	*−.183*	−.107	.010	−.117
o. Choices expected	.020	*.175*	−.137	.168	−.042	−.077	−.105	.028	−.022	.083	.044	.016	.103	−.008	*−.227*
p. Choices received (popularity)	.156	−.075	.002	−.061	.042	.070	−.119	.119	−.048	.008	.061	.059	−.126	.152	−.046
q. Expectancy he will choose	.014	−.054	.157	*−.192*	.032	.090	.020	.101	−.065	−.060	.008	−.090	.158	−.026	−.046
r. Age	.014	−.012	*.214*	.004	*−.002*	034	.035	*.176*	−.074	−.167	.167	.135	−.021	−.066	.142
s. Education	*.253*	−.062	.044	−.024	.028	.077	.171	−.061	−.044	−.014	−.053	.045	.066	−.038	−.017
t. Present grade	.109	.016	.125	.068	.047	.089	−.040	.107	−.009	−.102	.128	.142	.022	−.018	.086
Rate of initiation in interaction:															
u. 1. Showing solidarity	.032	−.084	−.006	.096	.081	*.258*	−.105	.062	−.017	.027	−.101	.073	−.027	−.032	.055
v. 2. Showing tension release	.050	−.065	−.084	.126	.051	.111	−.009	−.083	.026	.087	−.108	−.025	*−.218*	.065	−.050
w. 3. Showing agreement	.157	.022	.117	.081	.008	*.193*	.087	.162	−.105	−.158	.099	.124	.105	−.095	.054
x. 4. Giving suggestion	*.215*	−.125	.109	.032	.073	*.261*	.096	.114	−.142	−.060	−.105	*.223*	−.035	−.058	−.013
y. 5. Giving opinion	.027	−.059	−.058	−.023	−.014	.042	.035	.038	−.009	−.078	.092	.043	−.059	.017	.151
z. 6. Giving information	.008	−.012	−.004	−.052	.010	−.017	.072	−.027	.021	−.059	.023	.031	.058	−.059	−.006
aa. 7. Asking for information	*.251*	−.126	−.100	−.074	.125	.049	.079	−.053	−.054	.067	−.146	−.003	−.033	.051	.070
bb. 8. Asking for opinion	.142	−.013	.062	.037	.075	.073	.131	.080	−.096	−.119	.014	.165	−.033	.001	.020
cc. 9. Asking for suggestion	.001	−.001	.021	−.004	.014	.019	.012	.009	−.005	−.017	.011	.022	.012	−.010	−.014
dd. 10. Showing disagreement	.124	−.111	.012	−.012	*.193*	*.213*	.087	.005	−.025	−.052	−.060	.084	−.156	.076	−.077
ee. 11. Showing tension	.039	−.062	−.006	−.122	.002	−.096	.063	−.074	.013	.030	.027	−.116	−.036	.072	−.013
ff. 12. Showing antagonism	.069	.017	−.001	*.243*	−.039	−.011	−.112	−.002	−.027	.151	−.018	.054	*−.188*	.041	−.138
gg. Total rate	.138	−.088	−.011	−.017	.052	.121	.114	.026	−.037	−.093	.009	.101	−.027	−.029	.073
hh. Positive affect (1+2+3)	.137	−.034	.036	.134	.045	*.228*	.042	.079	−.063	−.057	−.009	.111	−.048	−.040	.023
ii. Instrumental action (4+5+6)	.047	−.047	−.015	−.039	.008	.046	.073	.014	−.008	−.083	.044	.070	.004	−.036	.061
jj. Instrumental action (7+8+9)	*.242*	−.116	−.045	−.049	.130	.075	.112	−.010	−.076	.003	−.102	.066	−.029	.038	.053
kk. Negative affect (10+11+12)	.081	−.091	−.003	.096	.057	−.026	.075	−.068	.001	.025	−.089	−.078	−.098	−.095	−.048

Table 2

Sign Test of Hypotheses

Hypothesis	Source	Criteria	Expected signs	Actual signs	Judgment
1. High *R* is associated with high intelligence	2, p. 53; 7, p. 208	f; g; h; i; j; k	+	+	not rejected
2. High *R* is associated with compulsiveness, drive for thoroughness	3, p. 72	gg, ee	++	++	in right direction
3. Subjects in good humor tend to have a high *R*	9, p. 21	v	+	+	in right direction
4. Sullen and unobliging subjects tend to have low *R*	9, p. 21	dd, ee	++	++	in right direction
5. High *R* is associated with manic conditions (depression hypothesis is conceived as identical here)	9, p. 21	hh, gg	++	++	in right direction
6. High *W* is associated with high intelligence	3, p. 20; 2, p. 10	f; g; h; i; j; k	+	+	in right direction
7. High *W* is associated with self-assertiveness	9, p. 59	gg	+	+	in right direction
		m	+	+	in right direction
8. High *W* is associated with compulsiveness	7, p. 245	ee, gg	++	−+	in wrong direction
9. High *W* is associated with manic conditions	2, p. 13; 9, p. 40; 3, p. 23	hh, gg	++	++	in right direction
10. Low *W* is associated with anxiety	3, p. 23	v, ee	−−	−−	in right direction
11. High *W* is associated with "overall thinking" (abstract ideas and major issues)	2, p. 14; 8, p. 57	x, y	++	++	in right direction
		h	+	+	in right direction
12. High *D* is associated with common sense or practicality	8, p. 58; 7, p. 261; 2, p. 14	a, b	++	−−	in wrong direction
13. High *D* is associated with depression	2, p. 15	hh, gg	−−	−−	in right direction
14. High *Dd* is associated with low intelligence	3, p. 28	f; g; h; i; j; k	−	−	not rejected
15. High *Dd* is associated with compulsiveness	2, p. 14	ee, gg	++	+−	in wrong direction
16. High *Dd* is associated with nagging and fault finding	3, p. 28	ff, x	++	+−	in wrong direction
		dd	+	−	in wrong direction
17. High *Dd* is associated with low adaptability	3, p. 27	a, b	−−	−−	in right direction
18. High *S* is associated with oppositional behavior	2, p. 47; 7, p. 266; 8, p. 60	dd, ff	++	+−	in wrong direction
19. Low *S* is associated with passivity	2, p. 47	w, ee	−−	++	in wrong direction
20. Total *Z* varies directly with intelligence	2, p. 12	f; g; h; i; j; k	+	+	not rejected
21. High *Z* is associated with elation	2, p. 13	v	+	−	in wrong direction
22. High *F*+ percentage is associated with attention to task and ability to concentrate	9, p. 56	x	+	−	in wrong direction
23. High *F*+ percentage is associated with depression	2, p. 19	hh, gg	−−	−−	in right direction
24. Low *F*+ percentage is associated with elation	2, p. 19	v	−	−	in right direction
25. High *F*+ percentage is associated with intelligence	2, p. 19; 3, p. 35, 37	f; g; h; i; j; k	+	+	in right direction

Table 2—*Continued*

Hypothesis	Source	Criteria	Expected signs	Actual signs	Judgment
26. High *M* is associated with high intelligence	9, p. 26; 1, p. 268; 2, p. 23	f; g; h; i; j; k	+	+	in right direction
27. Low *M* is associated with depression	9, p. 27	hh, gg	++	+−	in wrong direction
28. High *M* is associated with elation	9, p. 27	v	+	+	in right direction
29. High color is associated with allocentricity and awareness of environment	2, p. 31	bb, cc, x	+++	+++	in right direction
30. High color is associated with excitability	9, p. 98	ee, ff	++	−−	in wrong direction
31. High color is associated with energy and initiative (with which subject attacks problems)	2, p. 31	x	+	+	in right direction
		a	+	−	in wrong direction
		l	+	+	in right direction
32. High color is associated with high affectivity	9, p. 76	hh, kk	++	+−	in wrong direction
33. Low color is associated with depression	2, p. 35; 9, p. 31; 3, p. 45	hh, gg	++	−+	in wrong direction
34. High shading is associated with depression	2, p. 35, 36	hh, gg	−−	++	rejected
35. High shading is associated with passivity	2, p. 35	w, ee	++	+−	in wrong direction
36. High shading is associated with anxiety	3, p. 51	v, ee	++	+−	in wrong direction
37. High vista is associated with the feeling of inferiority	2, p. 33; 9, p. 201	ee	+	+	in right direction
38. High vista is associated with depression	2, p. 34; 9, p. 201	hh, gg	−−	++	in wrong direction
39. Animal form is inversely related with intelligence	2, p. 15	f; g; h; i; j; k	−	+?	in wrong direction
40. High animal response is associated with childhood	8, p. 80	r	−	−	in right direction
41. High anatomy responses are associated with hypochondriasis	2, p. 43; 3, p. 63, 64	ee	+	−	in wrong direction
42. High variety of content is related to high intelligence and high education	7, p. 272; 2, p. 41	f; g; h; i; j; k	+	−?	in wrong direction (but word fluency significant)
43. High variety of content is related to high education	7, p. 272; 2, p. 41, 42	s	+	+	in right direction
44. Low number of popular responses is associated with low intelligence	2, p. 17	f; g; h; i; j; k	+	+	in right direction

terpretation of protocols, these results indicate that if the generalizations implicit or explicit in the Rorschach theory are jointly considered, except in the case of total *R*, the interpreter will be on quite tenuous ground. The ordering of relationships between the Rorschach variables and the criteria is not dissimilar from chance expectation. Further, not only are most of the relationships of zero order, but the only ones in which this is not the case involve intelligence, which can be approached more efficiently by other instruments.

The implications of findings of this order are twofold: (*a*) While one might question the validity of the criteria, the questioner needs to provide better criteria to demonstrate the relationships exist in fact. Pointing to the significant differences found in carefully preclassified groups is not a demonstration of the validity of the measure. It does not constitute forward selection or prediction. Equally im-

portant for consideration is that the preselection criteria may be easier to apply than the test; *extreme* behavior may be recognized by other tools (including even direct questions). (*b*) The matrix of intercorrelations indicates that, although the amount of variance under control in any given instance is small, for limited purposes a forward validation study of the Rorschach may indicate utility of some of the arbitrary scoring categories in use.

Received September 8, 1954.

References

1. Bales, R. F. *Interaction process analysis.* Cambridge, Mass.: Addison-Wesley, 1950.
2. Beck, S. J. *Rorschach's test.* Vol. II. New York: Grune & Stratton, 1949.
3. Bochner, Ruth, & Halpern, Florence. *The clinical application of the Rorschach test.* (2nd Ed.) New York: Grune & Stratton, 1945.
4. Borgatta, E. F., & Bales, R. F. Task and accumulation of experience as factors in the interaction of small groups. *Sociometry,* 1953, **16**, 239–252.
5. Eschenbach, A. E., & Borgatta, E. F. *A review of some major Rorschach scoring categories and hypothesized personality correlates. USAF Hum. Resour. Res. Inst. Res. Mem.,* 1953, No. 19.
6. Hughes, R. M. Factor analysis of Rorschach diagnostic signs. *J. gen. Psychol.,* 1950, **43**, 85–103.
7. Klopfer, B., & Kelley, D. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1946.
8. Mons, W. *Principles and practice of the Rorschach personality test.* London: Faber & Faber, 1948.
9. Rorschach, H. *Psychodiagnostics.* (4th Ed.) New York: Grune & Stratton, 1949.
10. Williams, H. L., & Lawrence, J. F. Further investigation of Rorschach determinants subjected to factor analysis. *J. consult. Psychol.,* 1953, **17**, 261–264.
11. Wittenborn, J. R. A factor analysis of Rorschach scoring categories. *J. consult. Psychol.,* 1950, **14**, 261–267.
12. Wittenborn, J. R. An evaluation of the use of different scores in prediction. *J. clin. Psychol.,* 1951, **7**, 108–111.
13. Wittenborn, J. R., & Holzberg, J. D. The Rorschach and descriptive diagnosis. *J. consult. Psychol.,* 1951, **15**, 460–463.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Predicting Success on the Member-Employee Rehabilitation Program[1]

Bernard A. Stotsky
VA Hospital, Brockton, Massachusetts

To aid in rehabilitating chronic hospitalized psychotics, the Member-Employee program was instituted in VA-NP hospitals. Selected patients were discharged and placed in full-time salaried positions as employees of the hospital. Employment in a normal work atmosphere was designed to help them make the transition from hospital to community more easily. Selection of member-employees by the Medical Rehabilitation Board was mainly on an intuitive basis.

This study was undertaken to discover objective criteria which could be used to predict success on the program. The Guilford-Zimmerman, Manson, Revised Beta, Purdue Pegboard tests, patient attitude, Palo Alto Hospital Adjustment, Northport rating scales, and information on seven biographical items were collected for the first 18 patients approved by the board. Of all the tests, scales, and items used, only three of the biographical items—*arrests for any except psychiatric reasons, difficulty with alcohol prior to the past two years,* and *present difficulties with alcohol*—differentiated the successful from the unsuccessful member-employees. With the sample increased to 36, differences for the three items were significant at .05 level. The item, *present difficulties with alcohol* classified 35 out of 36 correctly, being significantly related to failure ($p < .001$). Only this item predicted significantly better than the board ($p = .02$). For a second sample of 25 from another VA hospital, the item did not discriminate better than chance though the trend was positive ($p = .11$).

Two main reasons for failure on the program were exacerbation of psychotic symptoms and overindulgence in alcohol. From these findings it appears that alcoholics and incompletely stabilized psychotics are poor prospects for a program of this type.

Brief Report
Received March 9, 1955.

[1] An extended report of this study may be obtained without charge from Bernard A. Stotsky, 1233 Warren Ave., Brockton, Massachusetts, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4555 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Relations Among Several Objective Measures of Anxiety Under Different Conditions of Motivation[1]

Anthony Davids
Harvard University

Psychologists have long agreed that the problem of human anxiety is of great theoretical and practical significance, but it seems to have taken the development of a rather simple questionnaire to open the floodgates and release a tidal wave of experimental findings. We will not even attempt to review the host of publications that have appeared on the psychological scene since Taylor recently reported on her scale for measuring manifest anxiety (18, 19). Performance on this scale has been related to a wide variety of experimental measures, including such things as eyelid conditioning (9, 16, 17, 18), verbal learning (11, 15, 20), word associations (3), reactions to stress (4, 22), responses to projective tests (8, 22), and various other complex behaviors (1, 7, 14, 21). Also, a comprehensive review of these studies and a great many others has been prepared (13). So, rather than discuss studies that are probably familiar to anyone acquainted with the recent literature, we will turn immediately to consideration of the present experiment.

In the attempt to discover what unique advantages or disadvantages this new anxiety inventory might have, results were compared on the following instruments: (*a*) Taylor anxiety scale; (*b*) Psycho-somatic Inventory; and (*c*) anxiety self-ratings.

More specifically, there were two main purposes in this study. First, we wished to examine the intercorrelations between results on the different measures within each of two groups of subjects. Second, we wished to see what effect the experimental manipulation of motivation would have on responses to the various objective measures of anxiety.

Method

Subjects. There were two experimental groups, each consisting of 20 male undergraduates selected from an introductory psychology course. With the exception of the experimental manipulation of motivation, described below, the procedure for selecting members of both groups was designed to control as many relevant variables as possible. The groups were matched to the extent where we felt quite certain that any differences revealed on the experimental measures could probably be attributed to the different conditions of motivation under which they were examined.

Treatments. The first group of 20 *Ss* completed the questionnaire and self-ratings in the course of their service as paid volunteers participating in an intensive investigation of personality. They believed that the results were anonymous, and were informed that the data would be used solely for purposes of "scientific research" and would not affect them personally in any way. The second group of 20 *Ss* was comparable to the first, with the exception that they were informed the results would be used to select one or two members from the group to work as

[1] This study, conducted at the Harvard Psychological Clinic as part of a program of research under the direction of Professor Henry A. Murray, was facilitated by the Laboratory of Social Relations. The research project was supported in part by a grant from the Rockefeller Foundation, and by grant M–700 from the National Institute of Mental Health, Public Health Service. The author gratefully acknowledges the assistance of Mr. Leo McNamara, who helped with the scoring and statistical analysis of the data.

highly paid assistants to staff members conducting research on small group interaction. They were led to believe that we were looking for mature, well-adjusted individuals who would be able to observe and analyze the behavior of others with minimal interference from their own problems and abnormalities. At the time of selection, an interview was conducted with each *S* in order to ensure that only individuals with high motivation to secure the position would be used in the present study.

Measures. The Taylor scale (19) consists of 50 items selected from the Minnesota Multiphasic Personality Inventory (MMPI) on the basis of clinical judgment that the items are indicative of manifest anxiety. These statements, to which *S* answers either "true" or "false," are embedded in a Biographical Inventory which contains 175 additional buffer items that were also taken from the MMPI. The maximum possible anxiety score is 50 and the minimum is zero.

The Psycho-somatic Inventory (12) consists of 92 questions, of which 46 refer to psychological experiences and 46 refer to physiological experiences. These experiences are supposed to differentiate the "psychoneurotic" from the "normal" individual. The questions are answered on a 4-point scale running from "often" to "never," and the total score based on the entire inventory provides a measure of "neuroticism." High positive scores are indicative of normality and negative scores are indicative of neuroticism.

The anxiety self-rating scale (2) consists of five parts. There are four subratings in response to specifically defined experiences and feelings indicative of anxiety, and one general rating. The four subratings are concerned with (*a*) illness, (*b*) irrational fears, (*c*) insecurity, and (*d*) anxiety symptoms. For each subarea several examples are provided in order to leave little doubt as to the kinds of thoughts and behaviors being assessed. The general rating follows the four subratings, and the instructions read: "Comparing yourself with others of the same sex and age, rate your usual level of general anxiety by placing a cross on the appropriate point on the following scale." Each of the five separate ratings is made on a 6-point scale, running from "much less than most people," to "much more than most people." Summing scores for the five ratings, the maximum anxiety score obtainable is 30 and the minimum is 5.

Predictions. Our expectation was that there would be a significant positive association between the different measures of anxiety within each group, and that there would be less anxiety evidenced by the 'job seekers" than by the "helpers of science."

Table 1

Rank-Order Correlations Among Three Measures of Anxiety Under Two Conditions of Motivation *

Measure	Taylor anxiety scale		Psycho-somatic Inventory	
Anxiety self-ratings	*.60***	.74	*.74*	.84
Taylor anxiety scale			*.76*	.83

* Italicized coefficients are for the 20 "research volunteers" and nonitalicized coefficients are for the 20 "job seekers."
** A coefficient of .52 is significant at the .01 level for a one-tailed test.

Results and Discussion

The results presented in Table 1 show that within both groups there are significant intercorrelations among the three different measures of anxiety. In both groups, the correlation between the manifest anxiety scale and the Psycho-somatic Inventory is not much lower in magnitude than reliability coefficients reported for the manifest anxiety scale (19). These results indicate that the Taylor scale is not unique. Regardless of the particular instrument employed to assess anxiety, the relative standing of the *S*s is quite similar.

Comparing the mean anxiety scores for the two groups, it is evident from Table 2 that *S*s who are motivated to present a favorable picture of themselves reveal less anxiety in response to each of the three assessment methods. Rather unexpectedly, it is the self-rating scales that show the least difference between the two groups. In response to the Psychosomatic Inventory, members of the group with nothing to gain by deceiving us reveal themselves as more neurotic. That is, they admit having more experience with disturbing anx-

Table 2

Comparison of Anxiety Scores for Two Groups of 20 *Ss*

Measure	"Research volunteers"		"Job seekers"			
	Mean	Variance	Mean	Variance	*F*	*t*
Anxiety self-ratings	15.6	36.1	15.0	36.7	1.00	.29
Psycho-somatic Inventory	137.7	25730.9	202.2	10902.4	2.36*	1.50
Taylor anxiety scale	20.8	64.3	13.8	46.6	1.38	2.96**

* Significant at the .05 level.
** Significant at the .01 level.

iety (lower scores indicate greater disturbance) than do the *Ss* who are motivated to appear "well adjusted." The difference between these means does not attain the .05 level of significance, but Bartlett's test reveals that the variance for the group of volunteer research *Ss* is significantly greater, which is what we would expect in a group that is not motivated to distort their responses in the direction of social acceptability.

Although the above differences are in accord with our theoretical predictions, it is only the manifest anxiety scale that shows a statistically significant difference between the means for the two groups. This difference of seven points, with the highly motivated *Ss* showing less manifest anxiety, is significant beyond the .01 level, indicating that, in the present study, the Taylor scale is especially susceptible to *Ss*' conscious or unconscious motivation to present a favorable picture of themselves. This finding receives additional confirmation from the fact that seven *Ss* in the group of "job seekers" received manifest anxiety scores that were lower than the score of 10 achieved by the lowest-ranking *S* in the group of "research volunteers."

At this point, let us consider briefly the high degree of consistency among findings derived from different measures of anxiety. This is no isolated finding for several previous investigators have also found significant relations between anxiety measured by the Taylor scale and anxiety measured by other objective instruments. Holtzman, Calvin, and Bitterman (10) correlated Taylor anxiety scores with neuroticism measured by the Winne scale (23), and found a coefficient of .74. When the correlation was corrected for attenuation due to the unreliability of each scale, it increased to .86. These scales represent independent adaptations from the MMPI, and contain 7 overlapping items, but, nevertheless, the magnitude of this relationship is quite high. In another study, Deese, Lazarus, and Keenan (4) found a correlation of .69 between the Taylor scale and Winne scale.

The psychasthenia scale from the MMPI has also been found to relate positively to the Taylor scale. Deese, Lazarus, and Keenan (4) obtained a correlation of .81 between these scales, and a correlation of .40 after overlapping items had been eliminated. Eriksen (5) reported a correlation of .72 between these same scales. And Eriksen and Davids (6) secured a correlation of .92 between psychasthenia and manifest anxiety. This latter finding was reduced only slightly, to .86, when 8 overlapping items were excluded.

These results, coupled with findings in the present study, indicate that performance on the Taylor scale of manifest anxiety is highly related to (*a*) self-ratings on anxiety; (*b*) neuroticism on the Psycho-somatic Inventory; (*c*) neuroticism on the Winne scale; and (*d*) psychasthenia on the MMPI. In many instances, the magnitude of relationship between the independent scales is about as high as reliability coefficients for the Taylor scale. Thus, several seemingly diverse researches seem likely to be investigating very similar, if not identical, phenomena. It may be, as Eriksen and Davids suggest (6), that the Taylor scale is not so much a measure of amount of anxiety as it is a measure of the way people respond to or cope with their anxiety. At any rate, these consistent findings from independent investigations, relating results on the different objective measures to performance in a variety of psychological ex-

periments, suggest that these several lines of research should be brought together and viewed in some common framework.

In conclusion, we would like to point out that the Taylor scale is only one of many available objective techniques for assessing anxiety. And it may well be that other instruments, some of which have been established for years, provide even more valid measures of anxiety than does the Taylor scale. Evaluated according to the criterion of generating research, there is no doubt that this scale has made a notable contribution. Yet the reason why psychologists have seized upon this instrument with such eagerness is not easy to understand. In fact, one could ask, legitimately, "Why the emphasis on the Taylor scale?" An attempt to answer this question is beyond the scope of this paper. Here, we wish merely to stress the point that any one of several objective instruments could probably be employed with equal fruitfulness in the selection and classification of *Ss* for investigations of anxiety and behavior.

One thing, at least, seems certain. When objective measures are employed, and responses accepted at face value, the condition under which the *Ss* are examined is of considerable importance. When the examinees have something to gain by reporting fewer experiences with disturbing anxiety, these objective measures are quite susceptible to distortion. Results of the present experiment suggest that the Taylor scale may be even more susceptible to deception than are other objective methods for assessing anxiety.

Summary

Two groups of male undergraduates were administered the Taylor anxiety scale, Psycho-somatic Inventory, and anxiety self-rating scale. One group of 20 *Ss* took these procedures as volunteers participating in a research project, with anonymity guaranteed. The second group of 20 *Ss* was informed that the results of the test battery would be employed to select a few "mature, well-adjusted" individuals to work as paid assistants conducting psychological experiments.

In both groups highly significant intercorrelations were found among the three objective measures of anxiety. When the groups were compared, it was found that on the three measures the group of "research volunteers" achieved scores more indicative of manifest anxiety and maladjustment than did the group of "job seekers." Although these results were all in keeping with our theoretical expectations, it was only on the scale of manifest anxiety that the difference between the two groups attained the .01 level of statistical significance.

It was concluded that (*a*) the Taylor anxiety scale measures essentially the same variable or trait as that measured by several other available instruments, (*b*) the Taylor anxiety scale may be more susceptible to deception than are some of these other instruments, and (*c*) a number of seemingly heterogeneous researches may actually bear upon common phenomena.

Received November 4, 1954.

References

1. Bitterman, M. E., & Kniffin, C. W. Manifest anxiety and "perceptual defense." *J. abnorm. soc. Psychol.*, 1953, **48**, 248–252.
2. Davids, A. The influence of personality on auditory apperception and memory. Unpublished doctor's dissertation, Harvard Univer., 1953.
3. Davids, A., & Eriksen, C. W. The relation of manifest anxiety to association productivity and intellectual attainment. *J. consult. Psychol.*, 1955, **19**, 219–222.
4. Deese, J., Lazarus, R. S., & Keenan, J. Anxiety, anxiety reduction, and stress in learning. *J. exp. Psychol.*, 1953, **46**, 55–60.
5. Eriksen, C. W. Some personality correlates of stimulus generalization under stress. *J. abnorm. soc. Psychol.*, 1954, **49**, 561–565.
6. Eriksen, C. W., & Davids, A. The meaning and clinical validity of the Taylor anxiety scale and the hysteria-psychasthenia scales from the MMPI. *J. abnorm. soc. Psychol.*, 1955, **50**, 135–137.
7. Farber, J. E., & Spence, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–125.
8. Goodstein, L. D. Interrelationships among several measures of anxiety and hostility. *J. consult. Psychol.*, 1954, **18**, 35–39.
9. Hilgard, E. R., Jones, L. V., & Kaplan, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94–99.
10. Holtzman, W. H., Calvin, A. D., & Bitterman, M. E. New evidence for the validity of Taylor's manifest anxiety scale. *J. abnorm. soc. Psychol.*, 1952, **47**, 853–854.

11. Lucas, J. D. The interactive effects of anxiety, failure, and intraserial duplication. *Amer J. Psychol.*, 1952, **65**, 54–66.
12. McFarland, R. A., & Seitz, C. P. A psychosomatic inventory. *J. appl. Psychol.*, 1938, **22**, 327–329.
13. Mandler, G. Anxiety and performance: empirical and theoretical aspects of anxiety scales. Unpublished manuscript. Laboratory of Social Relations, Harvard Univer., 1954.
14. Matarazzo, J. D., Ulett, G. S., & Saslow, G. Human maze performance as a function of increasing levels of anxiety. *J. gen. Psychol.*, in press.
15. Montague, E. K. Role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91–96.
16. Spence, K. W., & Farber, J. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116–119.
17. Spence, K. W., & Taylor, Janet A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.
18. Taylor, Janet A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 91–92.
19. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
20. Taylor, Janet A., & Spence, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–66.
21. Wesley, Elizabeth L. Perseverative behavior in a concept-formation task as a function of manifest anxiety and rigidity. *J. abnorm. soc. Psychol.*, 1953, **48**, 129–134.
22. Westrope, Martha R. Relations among Rorschach indices, manifest anxiety, and performance under stress. *J. abnorm. soc. Psychol.*, 1953, **48**, 515–524.
23. Winne, J. F. A scale of neuroticism: an adaption of the Minnesota Multiphasic Personality Inventory. *J. clin. Psychol.*, 1951, **7**, 117–122.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

A Further Investigation of the Relationship Between Manifest Anxiety and Intelligence

Allen D. Calvin, Paul B. Koons, Jr., Joseph L. Bingham, and Howard H. Fink
Michigan State University

Grice (2), who worked with airman basic trainees, recently reported a significant negative relationship between Air Force Clerical Aptitude Index scores (which he states may be regarded as a general intellectual measure) and performance on the Taylor Manifest Anxiety Scale (*A* scale). Kerrick (3), who also worked with Air Force trainees, reported that the *A* scale correlates negatively with a number of Air Force tests of intellectual abilities. Matarazzo *et al.* (4) used college students as *S*s and failed to find a significant relationship between *A*-scale scores and performance on the CVS; however, they did find a significant negative correlation between *A*-scale scores and the ACE. Schulz and Calvin (5) failed in their attempt to replicate Matarazzo *et al.*'s finding of a significant negative relationship between *A*-scale scores and the ACE. Farber and Spence, who have made the most extensive contributions to the *A*-scale literature, state: "The authors have, in fact, been unable, over a period of years, to find any relation between *A*-scale scores of college students and conventional measures of intellectual ability, such as entrance-examination scores and grade-point averages" (1, p. 10).

Because of the widespread use of the *A* scale as a research and clinical instrument, a resolution of these contradictory findings seems called for. The present experiment investigates some of the factors leading to the equivocal results obtained in earlier studies.

Method

Subjects. The *S*s were obtained from two separate sources. The first group, hereafter referred to as Group A, were students in an undergraduate course in psychology at Michigan State University. The group consisted of 36 *S*s, 21 males and 15 females.

The other group, hereafter referred to as Group B, was obtained from a how-to-study course offered for students who were having scholastic difficulty. Two additional *S*s for this group were selected from a class in elementary psychology in which they were experiencing considerable difficulty with the course material. This group had an *N* of 15, 12 males and 3 females.

Procedure. The Wechsler-Bellevue Intelligence Test (7) was administered in individual sessions to each *S*. The *A* scale was presented with the Biographical Inventory items as suggested by Taylor (6).

Results

The Wechsler scores for Group A ranged from 102 to 141 with a mean of 122.75 and a σ of 9.18. This range, while not unusual for college students, is lacking in low scores; therefore we decided to obtain additional *S*s of lower intelligence, and Group B was obtained (range 92 to 129, mean 108.53, σ 10.41). However in order to obtain *S*s with such low scores it was necessary to select students having trouble with their academic work—this being an almost invariable concomitant of such scores in a university population. This, of course, creates some question as to the propriety of combining the scores of Group B with Group A. We decided to present the data for Group A alone and also to present the findings for the two groups combined. In interpreting the findings for the

Table 1

Pearson *r*'s between *A*-Scale Scores and Wechsler Performance

Item	Group A	Combined groups A and B
Full IQ	−.21	−.31*
Verbal IQ	−.22	−.39**
Performance IQ	−.15	−.18
Information	−.11	−.44**
Comprehension	−.19	−.25
Digit Span	−.37*	−.31*
Arithmetic	−.19	−.31*
Similarities	−.02	−.14
Vocabulary	.08	−.29*
Picture Arrangement	−.02	.13
Picture Completion	−.07	−.19
Block Design	−.36*	−.48**
Object Assembly	−.23	−.29*
Digit Symbol	.02	−.11

* Significant <5% level.
** Significant <1% level.

combined groups (range 92 to 141, mean 118.57, σ 11.56), the reader should, of course, bear in mind the manner in which Group B was selected.

The results for Group A alone and for the combined groups appear in Table 1. The correlations presented are between *A*-scale scores and the derived IQ scores (prorating for vocabulary where necessary), and between *A*-scale scores and raw subtest scores. Examining the findings for Group A first we see that the correlation between total IQ and manifest anxiety is not significant; however, an examination of the subtest correlations reveals significant negative correlations between *A*-scale scores and Digit Span and *A*-scale scores and Block Design.[1]

The results for Groups A and B combined indicate a significant negative relationship between *A*-scale scores and Full Scale IQ, Verbal IQ, Information, Digit Span, Arithmetic, Vocabulary, Block Design, and Object Assembly. Because of the previously mentioned scholastic difficulties of Group B which is confounded with their lower IQ scores, it is impossible to make any unequivocal interpretation of these results.[2] However, these findings indicate the necessity of restricting the conclusions drawn from a particular experiment to the population under study and the specific measure of intelligence used.

Discussion

Although we failed to obtain a significant negative relationship between *A*-scale scores and IQ in our "normal" group of students, which supports the previous findings of Schulz and Calvin (5) and Farber and Spence (1), our finding in the "normal" group of a significant negative correlation between *A*-scale scores and several Wechsler subtests, plus the highly significant results obtained when the students with lower IQ scores (and concomitant scholastic difficulties) were added, indicate the necessity for closely examining the intellectual make-up of the individuals used as *S*s in any experiment utilizing the *A* scale. It seems quite possible, as Kerrick (3) suggests, that the contradictory findings reported by various investigators who used the *A* scale to select high and low anxious groups for learning experiments, conditioning experiments, etc. were due to variations in the intellectual make-up of the respective experimental populations.

Summary

Two groups of college students were given the Wechsler-Bellevue Intelligence Test and the Taylor Anxiety Scale. Group A was composed of 36 students in an undergraduate psychology class. Group B was made up of 15 students with lower IQ scores who were having academic difficulty.

Significant negative correlations between *A*-

[1] It is interesting to note that these subtests have been mentioned by numerous clinicians as being good anxiety indicators, and some investigators using different measures of anxiety have also found these subtests able to differentiate anxious *S*s. The experimental findings are far from unequivocal; however, space precludes an adequate discussion of this topic here.

[2] Confounding of another type also exists in the relationship between anxiety and intelligence. As Grice (2) and Kerrick (3) pointed out, it may be that the highly anxious *S* is handicapped on an intelligence test and therefore obtains a lower score than his "true" score.

Our colleague Dr. Charles Hanley suggests another possibility. He feels that the *S*s of lower intelligence may have difficulty understanding the *A*-scale items and may inadvertently check items which raise their *A*-scale scores.

scale scores and several Wechsler subtests were reported for Group A, and when Group B was combined with Group A the *A*-scale scores of the combined groups showed a significant negative correlation with Total IQ, Verbal IQ and a large number of subtests. The implications of these results for contemporary research in this area are discussed.

Received May 9, 1955.
Early Publication.

References

1. Farber, I. E., & Spence, K. W. Main and interactive effects of several variables on reaction time. *U. S. Naval Res. Tech. Rep. No. 3,* Contract N9 onr-93802, State Univer. of Iowa, Mar. 1, 1955.
2. Grice, G. R. Discrimination reaction time as a function of anxiety and intelligence. *J. abnorm. soc. Psychol.,* 1955, **50**, 71–74.
3. Kerrick, J. S. Some correlates of the Taylor Manifest Anxiety Scale. *J. abnorm. soc. Psychol.,* 1955, **50**, 75–77.
4. Matarazzo, J. D., Ulett, G. A., Guze, S. B., & Saslow, G. The relationship between anxiety level and several measures of intelligence. *J. consult. Psychol.,* 1954, **18**, 201–205.
5. Schulz, R. E., & Calvin, A. D. A failure to replicate the finding of a negative correlation between manifest anxiety and ACE scores. *J. consult. Psychol.,* 1955, **19**, 223–224.
6. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.,* 1953, **48**, 285–290.
7. Wechsler, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

The Correlation of Intelligence and Neurological Findings on Twenty-Two Patients Autopsied for Brain Damage[1]

Robert S. Morrow
VA Hospital, Bronx, New York

and Joseph C. Mark [2]
VA Hospital, East Orange, New Jersey

This paper is the first part of a more extensive investigation of the effects of various types of structural brain pathology on the psychological test performance of twenty-two adult males. In each of these cases the brain pathology was the main cause of death as determined by autopsy. All of these former patients had received a routine battery of psychological diagnostic tests as well as the usual neurological examination and techniques, i.e., EEG, laboratory studies, etc. The plan of the investigation is to correlate the psychological test results with the neurological and post-mortem findings.

The present paper deals solely with the effect of structural brain pathology on intellectual functioning as measured by the Wechsler-Bellevue Intelligence Scale and specifically it will report on:

1. The relationship between structural brain pathology and total as well as subtest Wechsler-Bellevue scale functioning. That is, the Wechsler-Bellevue results of this obviously brain-damaged group will be analyzed and compared with a control group.

2. Comparing the test scores of those whose locus of damage is in the anterior region of the brain (pre-Rolandic) with those whose lesions are in the posterior regions (post-Rolandic).

3. Comparison between the effects of focal and diffuse brain pathology on the intelligence test.

4. Comparing the effect of dominant cerebral hemisphere lesions with nondominant ones.

Historical Background of the Problem

The literature on the relationship between brain disturbance and psychological functioning is extensive, even when confined to studies of patients rather than animals. Klebanoff (17, 18) has written two excellent surveys. The recent use of intelligence tests such as the Wechsler-Bellevue Intelligence Scale in the studies of brain functioning has yielded valuable information regarding some of the older theories of brain functioning, such as localization theories, and has raised new and important hypotheses. For the most part,

[1] A modified version of this paper was read at the American Psychological Association, New York, September, 1954.

[2] Formerly with the Clinical Psychology Section, Neuropsychiatric Service of the Veterans Administration Hospital, Bronx, New York where this work was done. The authors wish to express their indebtedness to the following: Dr. Louis Berlin, Assistant Professor of Neurology at Cornell University Medical College and former chief of the Bronx VA Hospital Neurology Service, for his assistance on the problems involving neurology, neuroanatomy, and neuropathology; and to the following Bronx VA Hospital staff members: Dr. Jacob Cohen, Assistant Chief Clinical Psychologist, for his help on the statistical evaluations; to Dr. H. L. Flowers, Chief of Neuropsychiatry, and Dr. E. G. Kreuger, Chief of Neurosurgery, for their cooperation and support, and to Dr. Harold G. Wolff, senior consultant in neurology whose searching questions and comments inspired this study.

however, the studies using the Wechsler-Bellevue test were concerned with the development of test patterns as diagnostic signs as indicated in the writings of Wechsler (27) and Rapaport (21). Allen's (2, 3, 4) studies are typical of this approach.

The major criticism of these studies has been their failure to use adequate control groups to permit the comparison of pathological brain function with the function of persons free from structural brain pathology. Many studies have been reported on the localization of intellectual functions (6, 8, 9, 10, 12, 13, 14, 19, 20, 22, 23, 25, 26) and these have been criticized for failing to use specific anatomical data. This study attempts to take care of both these important objections.

Subjects and Procedures

In this investigation, a brain-damaged group was compared with a control group of psychiatric patients. Each group contained 22 male patients from the Veterans Administration Hospital in the Bronx, New York. The death and post-mortem examination of the brain-damaged patients occurred within the past seven years during which time each had been referred at least once for psychological diagnostic examination from the neurological or neurosurgery services. They were routine referrals handled by different staff and trainee psychologists. Wherever possible each patient received Form II of the Wechsler-Bellevue Intelligence Scale, the Wechsler Memory Scale, the Rorschach, Bender Gestalt, and Figure Drawing tests and in certain instances special tests (i.e., Goldstein-Scheerer, aphasia tests, etc.).

The psychological, neurological, and post-mortem data in this study come from the retrospective examination of the permanent records of the patients. The psychological tests were administered to the brain-damaged patients from one week to twenty-five months prior to death, the median being seven months. The tests were given only when the patient was, according to the judgment of both neurologist and psychologist, in good enough physical and mental condition to be tested.

The same battery of tests was given to the control group from the psychiatric wards who were matched for age, education, and color. The results of this matching as given in Table 1 show that both groups were approximately 40 years old and had about 10 years of formal education. Individual matching for occupation was not considered necessary because most of our patients come from the unskilled occupational groups. The psychiatric patients were either neurotic or mildly schizophrenic but were not at the time of testing receiving any physiological therapy.

Of the group of 22 brain-damaged patients, 18 had direct involvement of the cortex; the remaining 4 consisted of 2 subtentorial (posterior fossa) and 2 third-ventricle tumor cases. Eleven of the 18 cortical cases were considered focal in that the lesion, regardless of size, was circumscribed; these were mostly tumors. The other 7 were considered diffuse in that most or all of the cerebrum was affected, resulting primarily from vascular, toxic, or infectious disease. Of the 18 cortical lesions, 4 were anterior (pre-Rolandic), while 14 were predominantly posterior (post-Rolandic and pre- and post-Rolandic). Also, of these 18, 8 were in the left hemisphere, 2 were in the right hemisphere, and 8 were bilateral. All patients had a history of right-handedness. In making these determinations we were assisted by a neurologist who worked from the autopsy and neurological reports. The autopsy reports contained the microscopic as well as gross brain examination findings of the pathologist.

Results

The Wechsler-Bellevue intelligence test scores for the brain-damaged and psychiatric control groups are presented in Table 1. In addition, the Wechsler-Bellevue scores for 45 normal patients were available and are presented in the last column for further comparison. These patients, tested as part of another study which required normal controls, were recuperating from minor surgery (i.e., appendectomy, tonsillectomy, etc.) and had no history of either brain damage or psychiatric difficulty. They could not be used in this study as normal controls because they had received only the intelligence test and also because they were younger than the

brain-damaged group. The data for this group are given for the benefit of those who desire some form of normal reference point other than the standardized norms.

As can be seen from Table 1, the brain-damaged group is significantly lower in Full Scale and Performance IQ, but not significantly lower in Verbal IQ. Most of the subtest scores, with the exception of Vocabulary, Information, Comprehension, and Digits Forward, are significantly lower for the brain-damaged groups. This is particularly true for the performance tasks. The brain-damaged patients fared worst on the Digit-Symbol test, with an average weighted score of 4.23 and four of the subjects making zero weighted scores. On the other hand, the Vocabulary score of the brain-damaged group was the same as the controls, indicating that the vocabulary function in brain damage was apparently unimpaired. On the whole, the low performance test scores together with low Digit Span, Arithmetic, and Similarities corroborated the psychometric test patterns of organic brain pathology as reported in the literature.

Table 2 compares the 11 focal cases with the 7 diffuse ones. In addition, the 4 cases are given in which the lesion did not involve the cortex directly, the 2 third-ventricle and 2 subtentorial tumors. Since there are wide discrepancies in the ages of these groups, the subtest scores are presented as IQ equivalents rather than as weighted scores. The conversion of the subtest scores into an IQ equivalent is achieved on the basis of the standardization tables in Wechsler (27). Because of the paucity of cases, significance tests could not be made.

There is seen to be very little difference between the groups of focal and diffuse lesions, although both groups showed generally lowered test scores resulting from the brain damage. In contrast, the 4 noncortical cases showed, on the whole, very little intelligence test loss when compared with the control group. In both the focal and diffuse groups the Performance IQ is much lower than the Verbal IQ. Among the verbal subtests Vocabulary, Information, and Comprehension did not seem to be adversely affected.

The significance of these findings is that the lowered intelligence resulted from the cortical damage itself and seemed to be independent of the nature of the damage. It mattered little whether the lesion was focal or diffuse when the size of the lesion was somewhat controlled by virtue of the fact that both groups had extensive damage resulting

Table 1

Wechsler-Bellevue Intelligence Scale Scores of Brain-Damaged and Psychiatric Control Patients

Variable	Mean score: Brain-damaged (N = 22)	Mean score: Psychiatric (N = 22)	t*	p	Mean score "normals" (N = 45)
Age	40.91	39.82	.305	—	30.1
Education	10.00	10.09	.075	—	
Wechsler-Bellevue					
Full Scale IQ	91.93	108.85	3.647	<.001	111.9
Verbal IQ	97.67	106.76	1.897	—	110.2
Performance IQ	87.49	109.67	4.804	<.001	111.6
Subtest weighted scores					
Information	9.41	10.86	1.571	—	11.5
Comprehension	9.86	11.00	1.356	—	11.2
Digit Span (total weighted)	6.68	9.14	2.081	<.05	11.1
Digits Forward (no. correct)	5.73	6.36	1.518	—	7.2
Digits Backward (no. correct)	3.91	5.04	2.211	<.05	5.6
Arithmetic	7.09	10.41	2.781	<.01	10.5
Similarities	7.64	9.82	2.390	<.05	11.1
Vocabulary	11.05	10.89	.173	—	11.2
Picture Arrangement	7.05	10.38	3.112	<.01	11.3
Picture Completion	7.32	10.27	3.292	<.01	12.2
Block Design	7.00	10.00	2.988	<.01	11.1
Object Assembly	7.05	10.36	2.971	<.01	10.8
Digit Symbol	4.23	9.14	4.610	<.001	10.4

* df = 35–45.

Table 2

Wechsler-Bellevue Scores for Cases of Focal, Diffuse, and Subtentorial and Third-Ventricle Lesions

Variable	Focal (N = 11)	Diffuse (N = 7)	Subtentorial and 3rd ventricle (N = 4)	F*	p
Age	37.73	50.71	30.50	5.71	<.05
Wechsler-Bellevue					
Full IQ	90.09	87.57	105.25	.56	—
Verbal IQ	95.73	95.71	106.50	.64	—
Performance IQ	85.64	82.28	103.00	1.82	—
Subtest IQ equivalents					
Information	97	98	105		
Comprehension	102	96	109		
Digit Span (Forward & Backward)	86	93	101		
Arithmetic	85	92	103		
Similarities	87	89	106		
Vocabulary	106	104	106		
Picture Arrangement	93	83	100		
Picture Completion	89	92	97		
Block Design	90	85	106		
Object Assembly	90	74	113		
Digit Symbol	70	84	94		

* df = 2/19.

Table 3

Comparison of Wechsler-Bellevue Scores Between Pre-Rolandic and Post-Rolandic Lesions

Variable	Pre-Rolandic ($N=4$)	Post-Rolandic ($N=14$)	t*	p
Age	39.75	43.64	.53	—
Wechsler-Bellevue				
Full IQ	100.25	85.92	1.62	—
Verbal IQ	105.00	93.07	1.27	—
Performance IQ	94.50	81.42	1.33	—
Subtest IQ equivalents				
Information	103	94		
Comprehension	106	100		
Digit Span (Forward & Backward)	96	81		
Arithmetic	100	87		
Similarities	92	82		
Vocabulary	110	103		
Picture Arrangement	101	89		
Picture Completion	92	86		
Block Design	100	83		
Object Assembly	91	83		
Digit Symbol	79	73		

* $df = 16$.

in death. We could find no typical psychometric test pattern for either the focal or diffuse lesions since both show, on the whole, the typical pattern of gross brain damage.

Table 3 compares lesions anterior to the Rolandic fissure with posterior lesions. Here, too, the subtest scores are given as IQ equivalents because of the age difference. On the whole the patients in the pre-Rolandic or frontal lobe group show only slight intellectual interference, with most of the functions falling within the average range or better, except for the low Digit Symbol score. The intelligence impairment is much greater and much more consistent for those lesions which are either completely or almost completely in the areas posterior to the Rolandic fissure, although the differences are not statistically significant. Here average or better scores were obtained only on Vocabulary, Comprehension, and Information. In both cases, again, the Performance IQ was much lower than the Verbal IQ. The psychometric test pattern for structural brain pathology was much more clearly revealed for the group of post-Rolandic lesions than for pre-Rolandic ones. While few frontal lobe cases were involved in this study, the fact that these four cases show only slight impairment raises questions about some of the special intellectual functions ascribed to the frontal lobes.

Direct comparison between lesions of the left or dominant and right (nondominant) hemisphere is precluded because only two lesions were clearly right-sided and these were located in the frontal lobes. Instead, Table 4 shows the comparison between eight cases of left-hemisphere damage with eight that are bilateral. The two right-hemisphere lesions are also shown but not statistically compared. Despite the lack of statistical significance in the findings, the left-hemisphere lesions showed general loss in Verbal and Performance IQ while the bilateral group showed lowered Performance IQ only. Vocabulary, Information, and Comprehension were relatively unaffected for both groups, but the main differences were in the low Digit Symbol, Digit Span, Arithmetic, and Similarities subtest scores for the dominant hemisphere group, and low Digit Symbol, Picture Arrangement, Block Design, and Object Assembly subtest scores for the bilateral group.

Discussion

The small number of cases in each category precludes broad conclusions or generalizations. We feel justified in making tentative interpretations, however, because of the rarity of the instances where precise localization is available through autopsy. The Wechsler-Bellevue test patterns in this study generally

Table 4

Comparison of Wechsler-Bellevue Scores Between Left-Hemisphere and Bilateral Lesions

Variable	Left hemisphere ($N=8$)	Bilateral ($N=8$)	t*	p	Right hemisphere ($N=2$)
Age	37.12	50.00	2.20	(.06	
Wechsler-Bellevue					
Full IQ	84.12	88.00	.521	—	113.5
Verbal IQ	87.12	99.00	1.591	—	117
Performance IQ	83.25	78.63	.559	—	109
Subtest IQ equivalents					
Information	92	100			115
Comprehension	96	100			120
Digit Span (Forward & Backward)	80	92			108
Arithmetic	73	96			117
Similarities	80	92			101
Vocabulary	102	105			132
Picture Arrangement	95	80			106
Picture Completion	90	91			115
Block Design	89	82			109
Object Assembly	89	75			121
Digit Symbol	71	79			103

* $df = 14$.

Note.—The data for the two right-hemisphere lesions are shown in the last column but since the number is too small, in addition to both lesions being in the frontal lobe, no comparisons are made.

confirm the test patterns of most of the other investigations (21, 27). Some of the slight differences are worthy of comment. For example, Allen's (2, 3, 4) studies of both brain-damaged and brain-diseased patients obtained scores slightly higher than those of this study, but he worked with patients who were presumably healthier since they were seen in an out-patient clinic. Aita and his group (1) compared Wechsler-Bellevue scores of their entire brain-injured population with those classified into mild, moderate, and severe cases. The scores of the severely injured group bear a very close resemblance to ours. It seems, therefore, that the severity or extent of the brain disturbance influences the Wechsler-Bellevue test pattern. The unimpaired Vocabulary score confirms its usefulness as an index of premorbid intellectual capacity as recommended by Babcock (7) and others (21, 27). Other factors which might be involved in the patterning are contingent upon the nature of the disturbance, such as the location of the lesion, etc.

Whether the lesion is focal or diffuse is apparently of very little importance insofar as the effect of severe brain disturbance on the Wechsler-Bellevue test results is concerned. What seems to be of importance is the cortex itself as evidenced by comparisons of cortical lesions with subtentorial and third-ventricle lesions. Our findings seem to be in substantial agreement with those of Anderson (5, 6) who, in comparing what he regards as "focalized" gunshot wounds in the dominant hemisphere with the nondominant ones, reported a higher Wechsler-Bellevue Deterioration Quotient for the dominant-hemisphere cases. He also indicated that cases with dominant-hemisphere damage show a greater loss of verbal abilities while the nondominant-hemisphere cases show a greater loss in performance. McFie and Piercy (19) found that patients with left-sided lesions were inferior on the Weigl Sorting Test to patients with nondominant, right-sided lesions, irrespective of location in the frontal, parietal, or occipital lobes. Milner found that in left temporal lobe lesions causing aphasia the test pattern ". . . is likely to be one of severe deficit on verbal tests and relatively mild deficit on nonverbal ones" (20, p. 55). In our study, six of the eight dominant-hemisphere lesions showed some form of aphasia yet both Performance and Verbal IQ's were low, the Performance mainly because of the Digit Symbol and the Verbal mainly because of the Digit Span, Arithmetic, and Similarities. Weisenberg and McBride (28) showed that right-hemisphere damage resulted in lowered performance tests. Milner indicated that right temporal lesions in particular, as well as right parieto-occipital lesions, interfere with visuomotor functioning, mostly in terms of the perception of spatial patterns. Since there were in our study no right-hemisphere lesions other than the two in the right frontal lobes, the left hemisphere was compared with bilateral lesions. The bilateral involvement was generally similar to the reported findings on right-sided lesions in that the visuomotor tests were lowest, these being Picture Arrangement, Block Design, Object Assembly, and Digit Symbol.

Since the left-hemisphere lesions in this investigation showed both verbal and performance involvement, the question of aphasia merits consideration. Six of the eight dominant-hemisphere cases were reported as having some kind of aphasia, while two of the eight bilateral cases were reported as showing dysarthria but no aphasia. The speech impairment did not seem to interfere directly with the test administration. The fact that the Vocabulary as well as Information and Comprehension scores are either unaffected or only slightly so seems to argue against the language disturbance alone causing the generally lowered intellectual functioning.

Research on cerebral localization has been concerned mainly with frontal lobe function, as in the work of Goldstein (10, 11, 12), Rylander (22, 23), and Halstead (13). Halstead, typically, found that intellectual impairment was greater in frontal lobe lesions than in nonfrontal ones, regardless of whether the lesion was unilateral or bilateral, left or right. Our findings disagree. He found, also, no evidence for mass action by getting no correlation between extent of lesion and degree of impairment. In all, the frontal lobes were considered most essential to biological intelligence with the gradient being from the frontal lobes.

In contrast several studies refute the claims of frontal lobe function. Hebb (14, 15, 16) in his critical review of the work on the frontal lobes, as well as his own evaluation of four cases of frontal lobectomy in which he found no frontal lobe signs, concluded that the evidence on which intellectual defects have been attributed to frontal lobe damage is very poor. Teuber and his co-workers (24, 25, 26) found that problem solving in patients with posterior lesions is at least as difficult as for patients with frontal lobe lesions. The Mt. Sinai group under Bender, consisting of Battersby and his co-workers found that the "intellectual changes noted in patients with tumors of the frontal, temporal, parieto-occipital or even subtentorial areas were about the same in quality, although less frequent in the last instance" (8, p. 705). They stated that "no significant differences between patients with pre- or post-Rolandic lesions could be found" and concluded that "intellectual deficits may be produced by lesions in any portion of the cerebral hemispheres, and not chiefly by frontal lobe involvement, as some authorities are still claiming" (8, p. 712). Aita and colleagues (1) compared the deterioration quotients for anterior and posterior brain-injured patients and found much more "deterioration" for the posterior areas. In our study there was a difference between the frontal lobes and the posterior ones with the greater defects being, in general, in the post-Rolandic areas.

In trying to explain the disagreements among the different studies one should attempt to go beyond the problem of undifferentiated brain disturbance and take into consideration the following factors: (*a*) the use of matched controls; (*b*) the extent of the lesion; (*c*) its location in terms of cortical versus subcortical, dominance versus non-dominance, and pre- versus post-Rolandic lesions; (*d*) the nature of the lesion, such as focal versus diffuse, tumors versus vascular disturbances, and disease versus injury; (*e*) the stage of the disease, i.e., chronic versus acute, etc.; as well as (*f*) the kinds of tests used. Other factors and functions which may be of importance in reconciling the discrepancies might be the roles of factors such as aphasia, personality, and visuomotor function. In other words, the total picture of brain damage should be considered.

Summary and Conclusions

The purpose of this paper was to determine the effect of structural brain pathology on intellectual activity as measured by the Wechsler-Bellevue Intelligence Scale, Form II. The results of 22 adult males whose death resulted from various forms of brain pathology were compared with a control group of 22 psychiatric male patients. The autopsy examination reports indicated the nature, extent, and location of the lesions. The following are the main findings:

1. There appears to be a typical Wechsler-Bellevue test pattern for gross structural pathology consisting of significantly lower scores in Digit Symbol, Block Design, Digit Span (especially Digits Backward), Arithmetic, and Similarities. In addition, the Full Scale IQ and the Performance IQ, including all of the Performance subtests, were significantly lower for the brain-damaged population. There were no significant differences involving the Vocabulary, Information, and Comprehension scores. The seemingly unaffected Vocabulary score justifies its use as an indicator of premorbid intelligence.

2. Within the brain-damaged population significant differences were not found between the focal and diffuse lesions, although both groups showed lowered scores. The two posterior fossa and two third-ventricle cases showed practically no intellectual impairment.

3. Under the conditions of this study there is no indication of the so-called frontal lobe signs. On the contrary, patients whose lesions were located anterior to the Rolandic fissure showed far less intellectual loss than those with lesions which were post-Rolandic. The post-Rolandic lesions showed the typical loss and psychometric patterning of gross structural pathology.

4. Comparison between the left- or dominant-hemisphere lesions and bilateral lesions indicated in general that the dominant-hemisphere lesions show significantly lower Verbal IQ (mostly Digit Span, Arithmetic, and Similarities) and Performance IQ (mainly because of Digit Symbol) while the bilateral lesions

show only lowered Performance scores, including Digit Symbol, Block Design, Object Assembly, and Picture Arrangement.

These results are not compatible with the traditional brain theories which emphasized the roles of the frontal lobes and dominant hemisphere in intellectual functioning.

Received November 5, 1954.

References

1. Aita, J. A., Armitage, S. G., Reitan, R. M., & Rabinowitz, A. The use of certain psychological tests in the evaluation of brain injury. *J. gen. Psychol.*, 1947, **37**, 25–44.
2. Allen, R. M. The test performance of the brain injured. *J. clin. Psychol.*, 1947, **3**, 225–230.
3. Allen, R. M. The test performance of the brain diseased. *J. clin. Psychol.*, 1948, **4**, 281–284.
4. Allen, R. M. A comparison of the test performances of the brain-injured and the brain-diseased. *Amer. J. Psychiat.*, 1949–50, **106**, 195–198.
5. Anderson, A. L. The effect of laterality localization of brain damage on the Wechsler-Bellevue indices of deterioration. *J. clin. Psychol.*, 1950, **6**, 191–194.
6. Anderson, A. L. The effect of laterality localization of focal brain lesions on the Wechsler-Bellevue subtests. *J. clin. Psychol.*, 1951, **7**, 149–153.
7. Babcock, Harriet. An experiment in the measure of mental deterioration. *Arch. Psychol. N. Y.*, 1930, **18**, No. 117.
8. Battersby, W. S., Krieger, H. P., Pollack, M., & Bender, M. B. Figure-ground discrimination and the "abstract attitude" in patients with cerebral neoplasms. *Arch. Neurol. Psychiat.*, 1953, **70**, 703–712.
9. Battersby, W. S., Teuber, H. L., & Bender, M. B. Problem-solving behavior in men with frontal or occipital brain injuries. *J. Psychol.*, 1953, **35**, 329–351.
10. Goldstein, K. The significance of the frontal lobes for mental performances. *J. Neurol. Psychopath.*, 1936, **17**, 27–40.
11. Goldstein, K. *After-effects of brain injuries in war*. New York: Grune & Stratton, 1942.
12. Goldstein, K. Mental changes due to frontal lobe damage. *J. Psychol.*, 1944, **17**, 187–208.
13. Halstead, W. C. *Brain and intelligence: A quantitative study of the frontal lobes*. Chicago: Univer. of Chicago Press, 1947.
14. Hebb, D. O. Intelligence in man after large removals of cerebral tissue; report of four left frontal lobe cases. *J. gen. Psychol.*, 1939, **21**, 73–87.
15. Hebb, D. O. Man's frontal lobes: a critical review. *Arch. Neurol. Psychiat.*, 1945, **54**, 10–24.
16. Hebb, D. O. *The organization of behavior*. New York: Wiley, 1949.
17. Klebanoff, S. G. Psychological changes in organic brain lesions and ablations. *Psychol. Bull.*, 1945, **42**, 585–623.
18. Klebanoff, S. G., Singer, J. L., & Wilensky, H. Psychological consequences of brain lesions and ablations. *Psychol. Bull.*, 1954, **51**, 1–41.
19. McFie, J., & Piercy, M. F. The relation of laterality of lesion to performance on Weigl's sorting test. *J. Ment. Sci.*, 1952, **98**, 299–305.
20. Milner, Brenda. Intellectual function of the temporal lobes. *Psychol. Bull.*, 1954, **51**, 42–62.
21. Rapaport, D., Gill, M., & Schafer, R. *Diagnostic psychological testing*. Vol. I. Chicago: Year Book Publishers, 1945.
22. Rylander, G. Mental changes after excision of cerebral tissue. A clinical study of 16 cases of resections in the parietal, temporal and occipital lobes. *Acta Psychiat. Neurol.*, 1943, Suppl. **25**, 1–81.
23. Rylander, G. Psychological tests and personality analyses before and after frontal lobotomy. *Acta Psychiat. Neurol.*, 1947, Suppl. **47**, 383–398.
24. Teuber, H. L. Some observations on the organization of higher functions after penetrative brain injury in man. In Milbank Memorial Fund (Ed.), *Biology of mental health and disease*. New York: Hoeber-Harper, 1952.
25. Teuber, H. L., Battersby, W. S., & Bender, M. B. Performance of complex visual tasks after cerebral lesions. *J. nerv. ment. Dis.*, 1951, **114**, 413–429.
26. Teuber, H. L., & Weinstein, S. Performance on a formboard-task after penetrating brain injury. *J. Psychol.*, 1954, **38**, 177–190.
27. Wechsler, D. *The measurement of adult intelligence*. (3rd Ed.) Baltimore: Williams & Wilkins, 1944.
28. Weisenberg, T., & McBride, K. E. *Aphasia: a clinical and psychological study*. New York: Commonwealth Fund, 1935.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

The Awareness of Hostile Feelings in Patients by Their Nurses[1,2]

Sol Charen
Catholic University of America

Subjects were 25 male veteran patients hospitalized for six months and longer with pulmonary tuberculosis. Mean age was 33 years with a range of 20 to 56 years, educational level was at the mean of 10.5 grades with a range of sixth grade to three years of college. Associated with the *Ss* were four ward nurses who had daily contact with them during this entire period of hospitalization.

Patients were given the Moldowsky Hostility Test (1) which consists of 25 items from the MMPI dealing with the presence of hostile feelings. The nurses rated the patients through the Factor *H* scale, designed to measure the feelings of hostility in people (2). The correlation between the nurses' scores on the Factor *H* scale and the patients' scores on the Moldowsky was insignificant, − .113.

The experiment was repeated when the patients were transferred to another ward. After three months' stay they again filled out the Moldowsky and the three nurses from this ward rated them with the Factor *H*. The correlation was again insignificant (− .017).

The correlation between the ratings of the first group of nurses and the second group was .579, significant at the .01 level. Mean scores and sigmas for the patients on the test-retest situation were almost identical, with the correlation between tests being .563. The conclusion which the above correlations between nurses' scores and patients' scores suggest is that the nurses seem to have been rating external behavior of the patients although the Factor *H* scale calls for a measure of internal feelings and attitudes. There is obviously better than chance ability on the part of the two groups of nurses to agree in their estimation of a patient's hostility, less however to really know how he feels in this area. Despite close association with the subjects neither group of nurses gives test evidence for empathy.

Brief Report
Received March 29, 1955.

[1] From the Veterans Administration Hospital, Baltimore, Maryland. The statements and conclusions reached in this article are those of the author and do not necessarily reflect the opinion or policy of the Veterans Administration.

[2] An extended report of this study may be obtained without charge from Sol Charen, 200 Rhode Island Ave., N.E., Washington 2, D. C., or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4554 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

References

1. Dinwiddie, F. W. An application of the principle of response generalization to the prediction of aggressive responses. Unpublished doctor's dissertation, Catholic Univer. of America, 1954.
2. Lorr, M., Rubenstein, E., & Jenkins, R. L. A factor analysis of personality ratings of outpatients in therapy. *J. abnorm. soc. Psychol.*, 1953, 48, 511–514.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Psychological Changes over a Three-Year Period Following Bilateral Prefrontal Lobotomy[1]

Isidor W. Scherer, John F. Winne,
VA Hospital, Northampton, Massachusetts
and Robert W. Baker
Clark University

In a previous publication (1), the authors reported an investigation of the changes in psychological tests of male schizophrenic patients during the first year following prefontal lobotomy. This paper represents an extension of that report to the third year following surgery.

As has been discussed at length elsewhere (1, pp. 1–2), the study was so designed as to take into account several criticisms which may be made of preceding studies of lobotomy: (*a*) failure to control time intervals between successive assessments; (*b*) failure to use control groups as a baseline in terms of which changes in the operated group must be evaluated; (*c*) failure to use a wide variety of psychological measures; and (*d*) failure to integrate test findings with all available clinical data.

Experimental Procedure

Subjects. Fifty white male schizophrenic patients at the Veterans Administration Hospital, Northampton, Massachusetts, have served as subjects during some phase of this three-year study. Of these patients, 28 were lobotomized, the operation consisting of a standard open Lyerly-Poppen approach (1, p. 6) while the remaining 22 constituted the control group.

[1] From the Veterans Administration, Northampton, Massachusetts. An abstract of this paper was presented to the American Psychological Association, New York, September, 1954. The authors wish to thank Dr. Frank Politzer and Dr. Robert Fortier for their assistance in administration of the test battery and Dr. Cesareo Peña and Dr. Arnold Trehub for critical review of the manuscript.

The subjects for the present study consisted of 20 lobotomized and 16 nonlobotomized patients, fairly well equated for age, education, length of hospitalization, diagnostic classification, and degree of cooperativeness. All patients had received electroconvulsive treatment, which was unsuccessful or only temporarily beneficial, and all had been selected as suitable candidates for lobotomy by the medical staff of the hospital.

Test materials. The experimental materials consisted of a battery of 27 tests from which 113 measures were derived. A description of the tests employed, and the measures derived from them, may be found elsewhere (1, pp. 6–10). Insofar as possible, the test battery was group administered to the operated patients about three weeks before, and two weeks, three months, one year, and three years after lobotomy. The control group was similarly tested at comparable time intervals. Tests with published manuals were scored according to the prescribed procedure; the evaluation of qualitative data was accomplished through group judgment of three clinical psychologists (not including the authors).

Analysis of data. Each of the 113 variables was examined in order to ascertain the statistical significance of the *net* change in the experimental group relative to the control group from testing to testing, using either *t* tests or net shift in proportion (for dichotomized data). The emphasis, statistically, was on determining the net change in the operated group over and above the change ob-

served in the control group for the particular time period under consideration.

Problem of the Study

Four major conclusions were derived from the analysis of the data obtained in the one-year study (1). These were: (*a*) There is a tendency toward decreased mental efficiency at one year, associated possibly with organic alteration; (*b*) ego boundaries are strengthened, although a downward trend in the direction of weakened ego boundaries was noted in a comparison of the results at three months and one year after operation; (*c*) sexual awareness and differentiation are increased throughout the entire period, reflecting concern regarding the role which the patient assumes in interpersonal relationships; and (*d*) there is an increase in the rate of action, possibly reflecting a diminution of inhibitory factors. Related to *d*, perhaps, are tendencies toward lack of inhibition in the moral-social field.

The problem in the present study was primarily to ascertain the course of the test measures relevant to these conclusions, with particular emphasis on the period from one to three years following operation. Secondly, the data derived from all of the 113 measures were examined in order to ascertain the changes *within* the experimental group and *within* the control group, as well as the *net change* in the experimental group, for the entire period from before lobotomy to three years after lobotomy.[2] Finally, qualitative material obtained from observation of the patients in several different types of activities and from psychiatrists' progress notes was examined in an effort to determine any additional dimensions of difference between lobotomized and nonlobotomized schizophrenics over the three-year period.

[2] Tables giving, for each measure, the number of cases, means, the standard error of difference, t, and p for the changes within the experimental and control groups and for the net change in the operated group during the period from before operation to three years after operation and for the period from one year after operation to three years after operation have been deposited with the American Documentation Institute. Order Document No. 4556 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $2.00 for microfilm or $3.75 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Results

Changes during the period from one to three years following lobotomy. First to be considered are changes in those variables related to the first conclusion from the earlier study.

> There is a predictable pattern of test performance following lobotomy. Performance on tests measuring mental efficiency and organicity (malfunction due to organic factors) drops sharply during the two-week postoperative period. As recovery from the operation takes place, test performance returns to the preoperative base line or rises slightly above it, reaching a peak, in our study, at three months postoperation. There then appears to be a trend toward decline, such that the performance in some tests at one year postoperation falls below the base line, although several tests remain above (1, p. 20).

In the period from one to three years after prefontal lobotomy, a definite pattern suggesting improvement is noted. Of the 29 measures selected as pertinent to this variable, 15 show a shift, at the .20 level of significance or better, in the direction of increased mental efficiency during this time period. There are *no* measures which show a statistically significant net change in the negative direction. This over-all result is significant at the .01 level, when tested with χ^2 as described elsewhere (1, p. 10). The measures, together with the associated p values, are presented in Table 1, where *H* indicates a change in score signifying better efficiency. Thus, for example, abstraction as measured by the score on the Object Sorting Test is improved from one to three years after operation ($p < .01$); memory is improved, as indicated by increased accuracy in Visual Memory ($p < .05$) and more correct recollections in delayed recall on the Babcock Memory Paragraph ($p < .01$). Thus, the tentative trend toward reduced efficiency which had been observed during the first year is not continued, and, in fact, appears to be reversed.

The second conclusion held that ego boundaries are relatively strengthened one year following lobotomy. However, examination of the changes between three months and one

Table 1

Direction of Net Change in Measures of Mental Efficiency from One to Three Years Following Lobotomy

Measure	Net change	p
Organic Signs		
Word Association		
Phrases (83)	H	.01
Close (84)	—	—
Personal (87)	—	—
Draw-a-Man		
Shading (69)	—	—
Accuracy (65)	H	.20
Heavy Lines (71)	H	.20
Visual Memory		
Heavy Lines (12)	H	.20
Abstraction		
Categorization (57)	H	.10
Similarities (51)	—	—
Shipley Hartford Rights (55)	—	—
Object Sorting (48)	H	.01
Motor		
Tweezer Dexterity (31)	H	.05
Downey: Total Time (37)	H	.20
Halstead: Right Hand (38)	—	—
Motor Memory		
Halstead		
Both Hands (40)	H	.05
Right-Left (41)	—	—
Positions (44)	H	.10
Memory		
Hard Pairs (9)	—	—
Visual Memory		
Accuracy (10)	H	.05
Memory Par		
Immed. Right (18)	—	—
Delay Right (21)	H	.01
Total Right (24)	—	—
Immed. Distortions (19)	H	.20
Distorted Order (23)	H	.20
Digit Span		
Weighted Score (5)	—	—
Cognitive		
Digit Symbol		
Errors (1)	—	—
Weighted Score (2)	H	.05
Trailmaking		
Errors (59)	—	—
Serial Seven		
Errors (7)	—	—

Note.—The number in parentheses refers to the measure as defined in the one-year report (1, pp. 6–10). *H* indicates higher mental efficiency, — indicates no change in mental efficiency.

year after operation, although somewhat unclear, suggested that a trend in the direction of weakened ego boundaries had set in. Table 2, which lists the relevant measures, again demonstrates a reversal of trend during the period from one to three years after operation. Of 21 measures studied, 8 are seen to shift, at the .20 level of confidence or better, in the direction of strengthened ego boundaries. One measure, the number of omitted items on Word Association, moves contrary to the general pattern, with an associated p value between .10 and .20. This

Table 2

Direction of Net Change in Measures of Strength of Ego Boundaries from One to Three Years Following Lobotomy

Measure	Net change	p
Memory		
Digit Span		
Forward (3)	—	—
Memory Par		
Immed. Right (18)	—	—
Delay Right (21)	H	.01
Total Right (24)	—	—
Halstead		
Positions (44)	H	.10
Cognitive		
Digit Symbol		
Weighted Score (2)	H	.05
Halstead		
Right-Left (41)	—	—
Right-Both (42)	—	—
Vocabulary (17)	—	—
Body Image		
Draw-a-Man		
Accuracy (65)	H	.20
Contact Measures		
Rorschach		
Rejections (126)	—	—
Multiple Choice Rorschach		
Rejections (138)	—	—
Crude C responses (141)	—	—
Szondi		
p positive (105)	—	—
p negative (106)	—	—
Word Association		
Personal (87)	—	—
Omits (82)	L	.20
Tweezer Dexterity		
Lev of Asp (30)	H	.20
Attitude to God		
Positive (114)	H	.05
Negative (115)	H	.01
Aversion		
Staying at Home (118)	H	.20

Note.—The number in parentheses refers to the measure as defined in the one-year report (1, pp. 6–10). *H* indicates increased, *L* indicates decreased, — indicates no change in strength of ego boundaries.

Table 3

Direction of Net Change in Measures of Sexual Awareness and Concern from One to Three Years Following Lobotomy

Measure	Net change	p
Bender Gestalt		
Angle Difficulty (77)	H	.05
Draw-a-Man		
Sexual Confusion (66)	—	—
Accuracy (65)	H	.20
Nose Size (64)	—	—
Draw-a-Woman		
Shift in Breast Size (61b)	H	.20
Rorschach		
Reaction Time, Card VI (132)	H	.01
Aversion		
Masturbation (122)	H	.05
Incest (119)	H	.01
Deviant Sexual Behavior (120)	H	.10

Note.—The number in parentheses refers to the measure as defined in the one-year report (1, pp. 6–10). *H* indicates increased, *L* indicates decreased, — indicates no change in sexual awareness and concern.

over-all result is significant at the .01 level. We see, in Table 2, that measures of memory and cognition improve (p varies between .10 and .01), that more positive ($p < .05$) and fewer negative ($p < .01$) attitudes toward God are expressed, and that level of aspiration is increased ($p > .10 < .20$).

The third conclusion held that there is an increased awareness of sexuality following lobotomy which appears to be maintained for at least three months. Although this increase appeared to be present at one year, the course during the last half of the first year after operation was somewhat equivocal. Table 3, which presents the nine measures pertinent to this variable, indicates that sexual awareness increases during the period from one to three years after lobotomy. The over-all result is presumed to be significant at the .01 level, or better. Seven of the nine measures shift in a positive direction at the .20 level of significance or better. For example, we find increased aversion toward masturbation, incest, and deviant sexual behavior (p varies from .01 to .10), increased reaction time to Rorschach Card VI ($p < .01$), and better delineation of the sex in drawings of human figures ($p > .10 < .20$). Again, there are *no* measures which shift in a direction contrary to the general pattern.

The fourth conclusion maintained that there is an increase in the rate of action in lobotomized patients during the first year following lobotomy—that is, the operated patients tended to act more hurriedly or impulsively than did their controls. Table 4, which lists the direction of net change in the 24 relevant variables, exhibits an interesting phenomenon. Changes in three out of four tests which are primarily motoric (Total Inhibition score of

Table 4

Direction of Net Change in Measures of Rate of Action and Impulsivity from One to Three Years Following Lobotomy

Measure	Net change	p
Motor		
Downey Inhibition		
Total Time (36)	H	.20
T_3-T_1 (37)	H	.10
Finger Dexterity (29)	—	—
Tweezer Dexterity (31)	H	.05
Cognitive		
Digit Symbol		
Weighted Score (2)	H	.05
Serial Seven		
Time (6)	—	—
Word Association		
Time (88)	—	—
Rorschach Reaction Times		
Card III (129)	L	.20
Card IV (130)	L	.05
Card V (131)	L	.20
Card VI (132)	L	.01
Card VII (133)	L	.05
Card VIII (134)	L	.20
Card IX (135)	L	.20
Card X (136)	L	.10
Aversion Items		
Murder (116)	—	—
Theft (117)	L	.20
Incest (119)	L	.01
Deviant Sexuality (120)	L	.20
Withdrawal (121)	—	—
Masturbation (122)	L	.05
Cannibalism (123)	—	—
Expansiveness		
Visual Memory		
Size (11)	L	.01
Draw-a-Woman		
Size (60)	L	.05

Note.—The number in parentheses refers to the measure as defined in the one-year report (1, pp. 6–10). *H* indicates increased, *L* indicates decreased, — indicates no change in rate of action or impulsivity.

the Downey Will-Temperament Test, the difference between the first and third trials of the Downey, and the O'Connor Tweezer Dexterity Test) suggest that the rate of motoric action continues to increase during the period from one to three years following operation. Perhaps related to this is the increased score in Digit Symbol, a written test which involves speed of motoric action. On the other hand, the data obtained from all Rorschach reaction times, two measures of expansiveness, and four out of seven measures of moral-social attitude suggest that lobotomized patients become less impulsive and less likely to act rapidly in tests involving ideation or imagination.

In summary to this point, we see that lobotomized patients show the following net changes in the period from one to three years after operation: (*a*) a general improvement in tests of mental efficiency; (*b*) a strengthening of ego boundaries; (*c*) a continued increase in awareness of sexuality; (*d*) an increased rate of motoric action; and (*e*) more inhibition or a lesser readiness to act rapidly on tests involving imagination or ideation.

Changes over the three-year period. This paper has been concerned thus far with changes attributable to lobotomy with reference to the period from one to three years following surgery. Next to be considered are the changes over the whole time span employed—i.e., from immediately before lobotomy to three years following lobotomy. Very briefly, tests of significance of the net change in the operated group indicate that only 4 of the 113 measures show impairment which is statistically significant at the .20 level or better: Digit Symbol Errors, Trailmaking Errors, Word Association Omits, and Aversion to Cannibalism. Twenty-three measures show significant net gain, while the remaining 86 measures show no change. Since these results differ from a chance distribution at the .01 level, it may be concluded that the operated patients either held their own or did better on psychological tests over the three-year period than did their controls.

Since the interpretation of many of the 113 measures may be open to question, a group of 40 clear-cut measures of functional efficiency were analyzed in detail.[3] Table 5 presents these measures, the direction of change *within* the experimental and control groups for each of the four testings after operation, and the *net change* in the experimental group over and above the change in the control group. All changes shown are statistically significant at the .20 level or better.[4]

The data presented in Table 5 suggest that the experimental group, except at the second testing, tends to have better efficiency than evidenced prior to lobotomy. For example, three years after lobotomy, gains are shown in 24 tests, no change in 12 tests, and deficit in only 4—increased errors on Digit Symbol ($p < .05$), increased errors on Trailmaking ($p < .10$), and increased confusion in sequence for immediate and delayed recall of the Babcock Memory Paragraph ($p > .10 < .20$). The control group, on the other hand, shows considerable gain when tested for the second time, probably as a result of practice, but by three years has returned essentially to the level attained at the beginning of the experiment.

The final section of the table indicates the relative loss in efficiency of the operated group two weeks after lobotomy, followed by a slow but steady rise over the remainder of the three-year period. We believe that the failure to improve in tests of functioning efficiency—in spite of repeated exposure to the test materials—on the part of the nonoperated group is one aspect of the regressive

[3] Tables giving, for each of the 40 measures, the number of cases, means, the standard error of difference, t, and p for the changes within the experimental groups and for the net change in the operated group for the periods (*a*) Pre-2 week post, (*b*) Pre-3 month post, (*c*) Pre-1 year post, and (*d*) Pre-3 year postoperation have been deposited with the American Documentation Institute. Order Document No. 4557 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[4] It will be noted that there is considerable overlap between the measures presented in Table 5 and those presented earlier in Table 1 and Table 2. Table 5 includes all measures adjudged to yield unequivocal interpretation while Table 1 and Table 2 include only those measures which showed significant net change at some period during the three years following lobotomy.

Table 5

Direction of Change from Preoperative Baseline in Forty Tests of Functioning Efficiency over a Three-Year Period

Measure	Experimental group				Control group				Net change			
	2 wk.	3 mo.	1 yr.	3 yr.	2 wk.	3 mo.	1 yr.	3 yr.	2 wk.	3 mo.	1 yr.	3 yr.
Digit Symbol												
Errors (1)	—	—	—	L	H	—	—	L	—	—	—	—
Score (2)	—	H	H	H	H	—	H	—	—	H	—	H
Digit Span												
Forward (3)	—	—	L	—	H	H	—	H	L	L	—	L
Reversed (4)	—	—	—	H	—	H	—	H	—	L	—	—
Weighted Score (5)	—	L	—	H	H	H	—	H	L	L	—	—
Serial Sevens												
Time (6)	—	H	H	H	—	H	—	—	—	—	—	H
Errors (7)	—	L	—	—	—	—	—	—	—	—	—	—
Hard Pairs (9)	—	H	H	H	—	—	H	H	—	—	H	—
Visual Memory Rights (10)	—	—	H	H	—	L	H	L	—	H	—	H
Vocabulary (17)	—	—	H	H	—	L	—	—	—	H	—	—
Memory Paragraph												
Immediate Rights (18)	H	H	—	H	H	H	L	—	H	H	H	—
Delay Rights (21)	—	—	H	H	—	—	H	—	—	—	—	H
Average Rights (24)	—	—	H	H	H	—	H	—	L	—	—	—
Immediate Distortions (19)	L	—	H	H	H	H	H	—	L	—	—	—
Delay Distortions (22)	—	—	H	H	—	—	H	—	—	—	—	H
Average Distortions (25)	—	—	H	H	H	—	H	—	L	—	—	—
Immed. Distorted Order (20)	—	—	L	L	—	—	L	L	—	—	—	—
Delay Distorted Order (23)	—	—	L	L	L	—	—	L	—	—	—	—
Finger Dexterity Time (29)	H	H	H	H	—	H	—	—	H	H	H	—
Tweezer Dexterity Time (31)	—	H	—	—	—	—	—	—	—	—	—	—
Downey Total Time (36)	L	L	—	—	H	—	—	H	L	L	—	—
Halstead												
Right Hand (38)	—	—	H	—	H	H	H	—	L	L	L	—
Left Hand (39)	—	H	H	H	H	H	H	H	—	—	—	—
Both Hands (40)	—	—	H	H	H	—	H	—	L	—	—	—
Shapes (43)	—	—	—	—	—	H	—	—	—	L	—	—
Position (44)	—	—	—	—	—	—	H	—	—	—	—	—
Object Sorting												
Score (48)	—	—	H	—	—	H	H	—	L	L	—	—
Confabulation (49)	—	—	—	—	H	—	—	H	—	—	—	—
Symbolism (50)	—	—	—	—	—	—	—	—	—	—	—	—
Similarities (51)	—	—	—	H	H	—	—	—	L	—	—	—
Series Completion (53)	—	H	H	H	H	—	—	—	L	H	—	H
Shipley Hartford (55)	H	H	H	H	H	H	—	H	—	—	H	H
Categorization (57)	L	—	—	H	—	—	—	—	L	—	—	H
Trailmaking												
Time (58)	—	H	H	—	H	—	H	H	—	—	—	—
Errors (59)	—	—	L	L	H	—	—	—	L	—	L	L
D-a-W Confusion (62)	H	—	H	H	H	—	—	—	—	—	—	—
D-a-M Accuracy (65)	—	H	—	H	—	L	—	—	—	H	—	H
Confusion (66)	H	—	H	H	—	L	L	L	—	H	H	H
Bender-Gest. Elaboration (75)	L	H	H	H	—	L	—	H	L	—	—	—
Word Association Recall (81)	—	—	—	—	—	—	—	L	—	—	—	H
Number of Measures showing												
Gain	5	13	19	24	18	13	12	10	2	10	7	12
No change	31	24	17	12	21	22	23	23	25	23	31	26
Loss	4	3	4	4	1	5	5	7	13	7	2	2

Note.—The number in parentheses refers to the measure as defined in the one-year report (1, pp. 6–10). *H* indicates increased, *L* indicates decreased, — indicates no change in functioning efficiency.

character of schizophrenia. Further study of the apparent differential in this aspect of the regressive process is now underway, utilizing data obtained from the two groups five years after lobotomy.

Additional differences between lobotomized and nonlobotomized schizophrenics. The results which have been presented thus far have been largely quantitative in nature and convey a somewhat favorable connotation for the lobotomized patient. However, qualitative study of the behavior of operated patients and controls during the testing situation, ward activity, and in group and activity therapies suggests that in certain respects the operated group is not doing as well as the nonoperated. Although lobotomized patients appear to be more cooperative, they exhibit a great deal of childishness, lack of motivation, lack of goal-oriented behavior, concreteness of thinking, and irresponsibility—modes of behavior which do not seem to be related directly to the numerous test measures associated with functioning mental efficiency. Moreover, examination of psychiatrists' progress notes does not indicate that there is greater *clinical* improvement in the lobotomized patients than in their controls.

Discussion

The final portion of this report will first present briefly some possible explanations for the findings of increased cooperativeness and efficiency in the lobotomized patients. The apparent discrepancy between test performance and clinical status which was noted in the previous section will then be considered.

First, it appears that the immediate reaction of patients to the operation includes features such as would be found in any individual suffering cerebral trauma. Immediately after lobotomy such typical organic manifestations as increased irritability, concreteness, confusion, temporary release of repression, and a temporary defensive regression were observed. It appears that some of these factors work together to establish a degree of contact with the environment which was absent prior to the operation. The increased contact allows opportunity for social interplay and results in a breakdown of prelobotomy tendencies toward isolation.

A second factor to which we may point for explanation is the commonly observed reduction, following lobotomy, of anxiety associated with pathological preoccupation. This diminution of anxiety seems to remove the need to create new symptoms or to introduce other defensive measures, and so releases psychic energy for employment in the solution of reality-related problems. This could account for increased efficiency on psychological tests.

Third, it is possible that the reduction in the strength and number of morbid symptoms is in part the result of some degree of disorganization or interruption of ingrained thought processes which created the pathological defenses. In addition, the pathological defenses per se become "shaken up" by the organic assault so that they do not operate as effectively. Consequently, the possible deleterious effect of these pathological defenses on functioning efficiency is reduced and, concomitantly, the opportunity is created for new modes of expression of psychic energy.

Despite the increased efficiency noted on psychological tests, qualitative study of the clinical behavior of lobotomized patients does not appear to show the same positive gains. In part, this may be due to lack of clear knowledge of what the tests measure. Although *memory, abstraction, contact,* etc. are defined operationally, there is no certainty about the relationship of the definitions to their corresponding clinical concepts. More importantly, there appears to be no reason why performance on psychological tests should parallel that behavior which is considered in judgments of clinical status, since the latter represent different facets of the total personality and may depend, for measurement, on different variables. Performance on tests seems to depend on intelligence, cooperativeness, and the amount of psychic energy available for use in the solution of the problems which are presented, while judgments of clinical status depend upon a host of variables, many of which are ill defined or in the realm of subjectivity.

Moreover, the apparent lack of gain in clinical status may arise because the lobotomized patient is making more numerous contacts with his environment and is likely to be

more demanding on the ward personnel. In a sense, symptoms which are ordinarily encapsulated are more likely to be released and expressed following the operation. Because of this, observers are more apt to notice the operated patient's psychotic ideation and behavior than those of his less demanding and less active control and, consequently, adverse clinical judgments may be made.

Summary

This study deals with the changes in psychological tests in a group of schizophrenic patients during the three years following lobotomy, with particular emphasis on the period from one to three years following operation. The subjects consisted of 20 male white lobotomized schizophrenics and 16 nonlobotomized controls, all fairly well equated for age, education, length of hospitalization, diagnostic classification, and degree of cooperativeness.

All lobotomized subjects were group tested about three weeks before, and two weeks, three months, one year, and three years following lobotomy. Controls were similarly tested at comparable time intervals. The test battery consisted of 27 psychological tests from which 113 relatively independent measures were derived. The data obtained from each of the 113 measures were examined to ascertain the significance of the *net change* in the experimental group in the period from one to three years following lobotomy. Secondly, statistical tests were made of the differences *within* the experimental group, *within* the control group, and of the *net change* in the experimental group for the period from prior to operation to three years after operation.

The results are as follows:

1. Lobotomized patients show the following net changes in the period from one to three years after operation: (*a*) a general improvement in tests of mental efficiency; (*b*) a strengthening of ego boundaries; (*c*) a continued increase in awareness of sexuality; (*d*) an increased rate of motoric action; and (*e*) more inhibition or a lesser readiness to act rapidly on tests involving imagination or ideation.

2. Lobotomized patients generally did better on psychological tests over the three-year period than did their controls.

3. Lobotomized patients, except at the testing two weeks after lobotomy, tend to have better efficiency than that evidenced prior to lobotomy.

4. Control patients show considerable gain on the second testing, but at the end of the three-year period have returned essentially to the level attained at the start of the experiment. It is suggested that this lack of improvement in spite of practice on the tests is one aspect of the regressive character of schizophrenia.

5. Qualitative analysis of the lobotomized patients' behavior and examination of psychiatrists' progress notes suggest that in some respects the operated group is not doing as well clinically as the nonoperated group.

6. It is suggested that the increased efficiency on psychological tests evidenced by the operated group is due to (*a*) increased contact, (*b*) lessened anxiety which releases psychic energy for use in the solution of reality-related problems, and (*c*) disorganization of the pathological defenses as a result of the organic assault.

7. It is suggested, also, that the discrepancy, in the operated group, between psychological efficiency and clinical behavior may be due to (*a*) ill-defined test and clinical measures and (*b*) release of formerly encapsulated symptomatology which may encourage adverse judgments by ward observers.

Received November 10, 1954.

Reference

1. Scherer, I. W., Winne, J. F., Clancy, D. D., & Baker, R. W. Psychological changes during the first year following prefrontal lobotomy. *Psychol. Monogr.*, 1953, **67**, No. 7 (Whole No. 357).

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Diagnosis of Organicity by Means of Spiral Aftereffect

A. Cooper Price
VA Center, Biloxi, Mississippi
and H. L. Deabler
VA Hospital, Gulfport, Mississippi

One of the persistent problems facing a clinical psychologist working in a hospital setting is that of having a reliable and valid test for determining organic brain damage. Recent reports on the more frequently used tests such as the Rorschach, Wechsler-Bellevue, and others tend to cast some doubt on their ability for consistent differentiation of organic from nonorganic cases (9).

In 1949, Freeman and Josey (1) published a study in which ability to perceive a certain spiral aftereffect by patients with memory impairment was investigated. They rotated an Archimedes spiral for a period of time after which rotation was stopped. Normal subjects experienced a visual negative afterimage or effect of either expansion or contraction of the spiral immediately after the rotation was stopped. This apparent movement aftereffect has been referred to elsewhere as gamma movement (3). Freeman and Josey reported that most of their subjects (*Ss*) with clinically judged memory impairment did not report or were unable to perceive this distinctive aftereffect.

The results of their study led the present authors to believe that this phenomenon might be utilized for development of a technique for detection of organic brain involvement. It was hypothesized that nonorganic patients would be able to perceive the distinctive aftereffect while organic patients, especially those with cortical involvement, would be unable to perceive it or if they did perceive it, their perception of it would be incomplete. It was decided that research should be conducted to verify this hypothesis and, if results were favorable, to develop this technique as an aid for diagnosis of organicity.

Method

Apparatus. An electric motor and stand commonly used for color mixing experiments was used to produce the necessary rotation. To the shaft, an opaque disk of white plastic 6 inches in diameter was attached. An Archimedes spiral of 920° or 2½ circuits about the center was constructed of dark colored plastic and was placed against the white disk background. The spiral was reversed for the effect of opposite rotation. The disk was turned at approximately 100 revolutions per minute. It was later found that a spiral could be rotated on a phonograph turntable of 78 rpm with identical results.

Procedure. Testing took about five minutes per *S* and was conducted only in the daytime when illumination was good. The *S* was seated approximately 8 feet from the apparatus. Standard instructions were given immediately preceding the test as follows: "This is a special eye test (so stated in order to get attention and cooperation from psychotic *Ss*). Look at the center (pointing) and don't take your eye away until I tell you to." Then the rotation was begun. Spiral A, giving normal *Ss* a negative aftereffect of expansion, was presented first in half of the cases for 30 seconds. This was ample time for normal and other *Ss*. While the disk was rotating *S* was asked "What does the line appear to be doing?" This was done to make sure *S* was attending and could report his experience. After 30 seconds, the disk was stopped and *S* was

then asked immediately, "Now what does the line appear to be doing?" The patient's answer was recorded. The routine was repeated with the spiral reversed (Spiral B), giving normal subjects a negative aftereffect of contraction. This was followed by a repetition of presentation of A, then B for a total of four trials of 30 seconds each. Each normal report, that is, seeing the negative aftereffect correctly, was scored 1. Each abnormal report, that is, failure to perceive the negative aftereffect, was scored 0. Reports of aftereffect in which no apparent change of dimension occurred but in which apparent motion forward or backward was reported, were given a score of ½. Also, a doubtful or equivocal report was scored ½. The *Ss* scored a total of 4, 3, 2, 1, or 0 depending on the degree of their perception of the aftereffect. In the case of normals the negative aftereffect was so immediate and vivid that it was scarcely necessary to conduct the usual inquiry for scoring. There was little ambiguity as to the presence or absence of the aftereffect with most *Ss*. In the interests of conservatism, to give *S* the benefit of a doubtful response, fractional or half scores were raised to whole scores in final computation.

Strict adherence to the procedure for the testing of *Ss* was exercised. It was believed that variations from the standard directions might have considerable bearing upon the validity of the results obtained. Precaution was taken not to suggest or structure in any way the nature of the perception being investigated.

In order to demonstrate the perceptual aftereffect to be cortical phenomenon and not retinal (2), several normal *Ss* were tested with one eye covered. At the end of the 30-second exposure period, the rotation was stopped. At that time the *Ss* were instructed to cover the eye which had been used to view the rotating spiral and to use the other to view the now motionless disk. It was found that the aftereffect occurred just the same.

Subjects. The VA Hospital at Gulfport, Mississippi and the VA Center at Biloxi, Mississippi, supplied the *Ss* for this study. Three groups were used:

1. Forty normal adult males employed at the institutions at the time of testing, with no known history of organic involvement.

2. Forty nonorganic psychiatric cases. The diagnoses of these were based on hospital staff action. The group was composed of schizophrenics, manic-depressives, and neurotics. Patients recently or currently undergoing electroconvulsive treatment were not used in the study, since it is known that this form of treatment may produce temporary organic effects (4).

3. One-hundred and twenty organic cases with known cortical involvement. Diagnoses of these cases were based on hospital staff action. This group included cases of CNS syphilis, cerebral vascular accidents, chronic brain syndrome associated with alcoholism, cerebral arteriosclerosis, post-traumatic encephalopathy, and advanced stages of Parkinson's disease.

Table 1

Aftereffect Score

Subject group	Perception				
	None	Partial			Complete
	0 Score	1	2	3	4 Score
Normal ($N=40$)	0 0%	1 2.5%	0 0%	2 5%	37 92.5%
Nonorganic psychiatric ($N=40$)	0 0%	0 0%	1 2.5%	1 2.5%	38 95%
Organic ($N=120$)	72 60%	13 10%	24 20%	9 8%	2 2%

Results

Scores of *Ss* in the three groups and percentage computations are given in Table 1.

The chi-square technique was used to determine the possibility of a statistically significant difference between the organic and the nonorganic cases. Chi square for these data was found to be significant at the .001 level of confidence.

Discussion

Functional psychiatric cases were presumed to be without organic brain damage. The test data confirm this hypothesis, cases in this

group being able to perceive the aftereffect to a marked degree. The normal and the functional psychiatric *Ss* showed no significant difference in their performance.

Standlee recently published an article (7) on the spiral aftereffect as an indication of memory impairment with conclusions differing from those of Freeman and Josey (1) and with implications from our study. He administered the spiral test to 16 normal and 25 psychotic *Ss* and found that all were able to perceive the aftereffect. He also gave these same *Ss* the Wechsler Memory Scale and found that their Memory Quotient scores ranged from below 48 to above 143, lower and upper limits of the scale, with low scores indicating poor memory ability and high scores indicating superior memory ability. He reasoned from this wide variation in memory ability scores and from the aftereffect report of each of his *Ss* that "Freeman and Josey's 'strong probability of correlation' between the visual phenomenon and memory ability is not confirmed" (7, p. 317). Standlee has, however, made a crucial error in his reading and interpretation of Freeman and Josey's study, for they specifically stated that they were attempting to deal with memory *impairment*, not with variations in general memory ability. A distinction exists and must be made between these two conditions. A given individual may have low memory ability and may not have suffered from any type of brain damage or memory impairment. One of our nonorganic psychiatric *Ss*, for instance, was a primary mental defective, and he was able to perceive the aftereffect completely.

In the second part of his study, Standlee administered the same tests to his psychotic *Ss* approximately eight hours after their receiving one electric shock. On the spiral retest, he found two reporting incomplete aftereffect perceptions with 23 reporting both expanding and contracting perceptions. In his retesting for evidence of memory impairment, he used the same form of the Wechsler Memory Scale as previously used and found the average MQ score for the group actually higher than before shock. On the basis of these data, he concluded that electric shock did not impair memory ability or the ability to experience the spiral aftereffect.

This second part of Standlee's study is also open to criticism. First, his results do show that electric shock had some effect on perception of the aftereffect since two of his postshock *Ss* reported incomplete perceptions, and it is open to question as to how many of the other twenty-three drew upon or utilized their previous experience with the test in giving their reports on retesting. Practice effect may have been operative here as well as elsewhere. Second, use of a single shock likely had minimal effects. Had more been used, results would likely have been more marked as has been shown by other studies (4, 8). Third, as he himself admits, the use of the same form of the Wechsler Memory Scale for evidence of impairment is faulty because of practice effects, thus leaving him with no adequate criterion of memory impairment. His study would have fared better had he used in his postshock testing Form II which is available and which was used effectively by Stone (8).

While we did not set out to verify the relationship between memory impairment and the spiral aftereffect, our findings and conclusions, for reasons stated above, are more in agreement with those present in the study of Freeman and Josey than those of Standlee. We are in agreement with Standlee's study insofar as we both found psychotic *Ss*—ours specifically diagnosed as functional or nonorganic—as being able to perceive the aftereffect. His study, however, did not deal adequately with the impairment factor which may be of organic or of functional origin. If impairment exists and if it is of organic origin, we predict difficulty or inability to perceive the spiral aftereffect.

In summary, our data for the organic cases contrasted significantly with the nonorganic in the direction of expectancy. When failure to perceive the aftereffect is found, it is almost certain that cortical involvement exists.

The explanation of our significant results in psychophysical terms has not yet been worked out or established. Some leads may be found in Saucer's article on processes of motion perception (5) and in Shapiro's investigation on perception of apparent motion (6). Saucer is inclined to believe that motion perception is derived from a higher-level integrating or

synthesizing process in the central nervous system, this integrating or synthesizing process appearing to be adversely affected or nonexistent in individuals with organic brain involvement. Shapiro believes that the perception of apparent motion is related to an "irradiation of excitatory effects" (6, p. 58) and that there is "an exaggeration of inhibitory effects in brain damaged patients" (6, p. 67) which weakens or destroys motion perception. While adequate explanation of the phenomenon remains to be hypothesized and verified, the results are such as to justify present clinical use of this technique in diagnosis of organicity.

Conclusion

Organic cases with cortical involvement can be differentiated from nonorganic with high degree of certainty by means of the spiral aftereffect technique.

Received November 19, 1954.

References

1. Freeman, E., & Josey, W. E. Quantitative visual index to memory impairment. *Arch. Neurol. Psychiat.*, 1949, **62**, 794–796.
2. Gibson, J. J. The visual perception of objective motion and subjective movement. *Psychol. Rev.*, 1954, **61**, 304–315.
3. Graham, C. H. Visual perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 868–920.
4. Kalinowsky, L. B. Organic psychotic syndromes occurring during electric convulsive therapy. *Arch. Neurol. Psychiat.*, 1945, **53**, 269–273.
5. Saucer, R. T. Processes of motion perception. *Science*, 1954, **120**, 806.
6. Shapiro, M. B. A preliminary investigation of the effects of continuous stimulation on the perception of "apparent motion." *Brit. J. Psychol.*, 1954, **45**, 58–67.
7. Standlee, L. S. The Archimedes negative aftereffect as an indication of memory impairment. *J. consult. Psychol.*, 1953, **17**, 317.
8. Stone, C. P. Losses and gains in cognitive functions as related to electro-convulsive shocks. *J. abnorm. soc. Psychol.*, 1947, **42**, 205–215.
9. Yates, A. J. The validity of some psychological tests of brain damage. *Psychol. Bull.*, 1954, **51**, 359–380.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

The Efficacy of Diagnostic Pattern Analysis with the Wechsler-Bellevue[1,2]

Jacob Cohen

VA Hospital, Montrose, New York and City College, New York

Since the publication of the revised manual for the Wechsler-Bellevue scales (5), and more particularly, since the publication of the Menninger group (4), clinicians have attempted to draw inferences about personality and clinical diagnosis from the intersubtest differences of Wechsler-Bellevue records. This practice, variously referred to as pattern analysis, scatter analysis, and profile analysis, has many enthusiastic proponents in the clinical field. It has developed, in a remarkably short time, an enormous research literature with results which have been characterized by Magaret as "negative, conflicting, or questionable" (3, p. 370).

The evidence for the validity of pattern analysis comes mainly from researches in which the scores of criterion groups of patients are compared with those of controls, subtest by subtest. Significant differences discovered by *t* test are harvested and are then offered as differentially diagnostic signs.

The evidence against the validity of pattern analysis comes from several kinds of investigations: (*a*) subtest-by-subtest comparisons like those cited above which have failed to produce significant differences; (*b*) applications of previously adduced patterns or signs on new samples with failure to substantiate them; (*c*) reworking of data previously offered as indicative of differentiating patterns to demonstrate that, in fact, no such patterns exist.

[1] From the Clinical Psychology Section, Neuropsychiatric Service, Veterans Administration Hospital, Bronx, New York.

[2] The author wishes to acknowledge the cooperation of Dr. H. L. Flowers, Chief Psychiatrist, and Dr. Robert S. Morrow, Chief Clinical Psychologist, of Bronx VA Hospital. He also appreciates the efforts of the seven clinical psychologists who served as subjects.

The approaches of the researches described to the problem under question are indirect. The presence of significant differences between groups on some subtests is no guarantee either that the matching of the groups has been complete enough to remove the effects of extraneous variables, or that these differences are large enough to lead to fruitful applications on individual cases. The absence of significant mean differences does not preclude the possibility that two obverse patterns, either of which is identifiable by a clinician as pertaining to a single diagnostic group, may not cancel each other out and lead to differences which fail of significance. Wittenborn's (6) demonstration with Rapaport's data that pattern correlations within diagnostic groups are no larger than those between diagnostic groups—in the present author's opinion the clearest and most definitive attack on the problem—still leaves a logical loophole in that the clinician may differently interpret the same pattern when the level of the constituent scores varies. In all probability, this loophole does not loom very large from a practical viewpoint. Still, Wittenborn's approach to the problem is an indirect one.

The Problem

When directly stated, the question which is basically at issue is, "Can experienced clinicians, using only the Wechsler-Bellevue subtest weighted scores and the verbal, performance, and full-scale IQ's, make nosological diagnoses of individuals which are significantly related to neuropsychiatric criterion diagnoses?" If they can, a second question of degree of agreement arises and is of practical importance.

Materials

In the course of another investigation (1, 2), Wechsler-Bellevue data (eleven subtest weighted scores and the three IQ's) for 300 successive admissions meeting research criteria were gathered. These data were made up of groups of 100 cases each of neuropsychiatrically diagnosed psychoneurotics, schizophrenics, and brain-damaged patients. These were all white male veterans, aged 20 to 40, with unambiguous diagnoses which classified a patient into one and only one of the above three nosological groupings. The three groups were equated for mean and standard deviation with regard to age, years of education, and full-scale IQ. The frequency of subcategories in each group and the means and standard deviations of full-scale IQ, age, and years of education are given in the report of the earlier investigation (2).

Method

Subjects. The subjects proper were seven clinical psychologists with from four to fourteen years of diagnostic experience with the Wechsler-Bellevue. Two of the subjects had published research on the Wechsler-Bellevue.

Procedure. The 300 patterns were randomly ordered and submitted to each psychologist individually with the request to classify each pattern as either psychoneurotic, schizophrenic, or brain-damaged. The subjects were informed that there were 100 of each but that they need not maintain this equality in their diagnoses.

For each psychologist contingency tables were set up as follows: (*a*) A 3×3 table

Table 1

Degree of Agreement Between Psychologist A's Wechsler-Bellevue Pattern Diagnoses and Neuropsychiatric Criterion Diagnoses for the Categories of Psychoneurotic, Schizophrenic, and Brain-Damaged

	Neuropsychiatric criterion			
A's diagnoses	Psycho-neurotic	Schizo-phrenic	Brain-damaged	Total
Psychoneurotic	56	39	28	123
Schizophrenic	24	33	29	86
Brain-damaged	20	28	43	91
Total	100	100	100	300

Note.—$\chi^2 = 20.11$, $df = 4$, $p < .001$, $C_o = .31$.

Table 2

Degree of Agreement Between Psychologist A's Wechsler-Bellevue Pattern Diagnoses and Neuropsychiatric Criterion Diagnoses in Discriminating Psychoneurotics

	Neuropsychiatric criterion		
A's diagnoses	Psycho-neurotic	Other	Total
Psychoneurotic	56	67	123
Other	44	133	177
Total	100	200	300

Note.—$\chi^2 = 13.95$, $df = 1$, $p < .001$, $\phi = .22$.

was set up to determine the over-all nonchance accuracy of diagnosis between the psychologist's diagnoses and the neuropsychiatric criterion diagnoses. This type of evaluation is illustrated for one of the psychologists (A) in Table 1.[3] (*b*) To determine the nonchance accuracy of psychologist's diagnoses with regard to each diagnostic group separately, 2×2 tables were set up which effected a correlation between psychologist's diagnoses and the criterion diagnoses for each group with the other two combined. Table 2 illustrates this type of correlation for the psychoneurotic group utilizing the diagnoses of Psychologist A. (*c*) To determine the degree of relationship between pattern and criterion diagnoses, where chi square was significant, the corrected contingency coefficient was computed for the 3×3 table and phi coefficients for the 2×2 tables.

Results

Table 3 gives the results of the chi-square tests for the seven psychologists in terms of chi-square values. The first column of the chi-square values pertains to the 3×3 tests for each psychologist. The second, third, and fourth columns refer to 2×2 tests for psychoneurotics versus other, schizophrenics versus other, and brain-damaged patients versus other, respectively.

[3] Psychologist A was selected for use as an example of how the chi-square tests were performed. His was the most successful performance, and the reader would be seriously misled if he thought him typical. On the contrary, his attempt at prediction illustrates the inadequacy from a practical viewpoint of even the best performance in this group of experienced clinicians.

Table 3

Chi-Square Values of Significance of Contingency Between Psychologist's Wechsler-Bellevue Pattern Diagnoses and Neuropsychiatric Criterion Diagnoses

Psychologist	3×3	2×2			Number "correct"
		Psycho-neurotics vs. other	Schizo-phrenics vs. other	Brain-damaged vs. other	
A	20.11**	13.95**	1.38	11.38**	132
B	21.18**†	2.11	1.98	14.52**	115
C	8.44	.36	2.33	5.13*	117
D	2.41	.52	.57	1.92	111
E	6.74	.19	.01	.19	107
F	1.20	.06	1.05	.07	94
G	4.54	.75	.03	1.47	102

* $.05 > p > .02$.
** $p < .001$.
† Reflects in part some incorrectness.

Table 3 immediately indicates that Psychologists D, E, F, and G failed to agree with the neuropsychiatric criterion to an extent greater than chance expectation. This holds for both the over-all (3×3) contingency as well as for the treatment of the three groups separately (2×2).

Psychologist A, with 132 "correct" diagnoses (chance expectancy is 100 "correct") showed extra-chance agreement with the neuropsychiatric criterion in the over-all analysis ($p < .001$). In addition, he discriminated psychoneurotics from others and brain-damaged from others significantly ($p < .001$).

Lest the reader be overly impressed by the magnitude of the chi-square value and its high order of significance, stress should be placed on its exact meaning. It indicates that one can be quite confident that the hypothesis of no association is false; it does *not* mean that the *degree* of association is strong. The contingency coefficient for the 3×3 test for Psychologist A, corrected for the number of categories, was .31. For the psychoneurotic vs. other and brain-damaged vs. other associations, A's phi coefficients were, respectively, .22 and .19. This obviously leaves much to be desired in the way of accuracy.

Psychologist B's results may be misleading. Although the over-all contingency is very significant ($p < .001$), part of this large chi-square value is attributable to the fact that the number of "correct" diagnoses for schizophrenics is *below* chance expectancy. This discrepancy adds to the value of the chi square but, of course, detracts from the over-all correctness of the diagnoses. This is evident from the fact that he has a total of only 115 "correct" diagnoses, which is not significantly better than chance.

This is a good point at which to emphasize the fact that chi square measures the degree of departure of a set of empirical frequencies from some hypothetical set, and does not distinguish the direction of departure. When the hypothetical set of frequencies reflects the hypothesis of a chance association (the null hypothesis), departures in *both* directions, i.e., both less than chance and in excess of chance, equally increase the chi-square value. It is therefore necessary, when using this method, to carefully scrutinize the contingency table cell by cell, before concluding that a significant chi square means better than chance prediction.

Psychologist B does identify brain-damaged patients with above chance success ($p < .001$), but with a small degree of relationship, i.e., phi = .22.

Psychologist C, with 117 "correct" diagnoses, does not attain better than chance agreement with the criterion in the over-all comparison. However, he, too, identifies brain-damaged patients with better than chance success ($p < .05$), but with a phi coefficient of only .13.

Discussion

If the theoretical question is posed, "Is there any detectable association between a neuropsychiatric diagnostic criterion and diagnoses made from patterns?" the answer must

be in the affirmative. If the results of the less ambiguous 2 × 2 tests are considered it can be seen that four of these are significant, three of them at the .001 level. This result is unlikely to be accounted for by chance alone. It is also unlikely to be a chance phenomenon that three of these four significant relationships occurred with regard to the brain-damaged group, or that of the seven psychologists, three could significantly discriminate brain-damaged patients from others, while only one succeeded with the psychoneurotics and none with the schizophrenics. The conclusion that some experienced psychologists can make diagnoses from Wechsler-Bellevue patterns significantly better than chance is inescapable. It is also likely that the success they enjoy comes largely from the ability to discriminate significantly brain-damaged patients from others. This is the positive side of the picture.

If a more practical question were posed, namely, "How useful are Wechsler-Bellevue patterns for neuropsychiatric diagnosis?" the answer, on the basis of the present results, must be, "Not very." Four of the seven psychologists did not fare better than chance; indeed, one of them fell below mean chance expectancy. The most successful subject, A, made only 132 "correct" diagnoses out of the 300, and thus was correct 43 per cent of the time, not quite 10 per cent more than the 33⅓ per cent expected by chance. If due allowance is made for the 100 "correct" chance diagnoses, then this psychologist made 32 nonchance "correct" diagnoses for the remaining 200 cases and can be credited with correctly diagnosing 16 per cent of the cases on a nonchance basis.

The small degree of predictive accuracy is also apparent from the magnitude of measures of relationship employed. The best of these, a corrected contingency coefficient of .31 for A's 3 × 3 contingency, presents problems in interpretation, but is in any event quite low. The phi coefficient, which is a product-moment coefficient and hence interpretable approximately as would be a Pearson *r*, gave values ranging from .13 to .22 in the four instances of better-than-chance association.

That Wechsler-Bellevue pattern analysis cannot serve as a sufficiently valid independent diagnostic technique is so obvious that it is not worthy of comment. The data suggest, however, that even as a supplementary diagnostic aid it is of questionable value. For many psychologists (four out of the seven well-qualified subjects of the present investigation), impressions derived from the pattern should be given no weight in the diagnostic decision. For some, it should play so small a role as to be of little consequence.

Summary

Seven experienced clinical psychologists attempted to diagnose the Wechsler-Bellevue patterns of 300 neuropsychiatrically diagnosed veteran male patients as psychoneurotic, schizophrenic, or brain-damaged. Using neuropsychiatric diagnoses as criteria, it was found that only one clinical psychologist correctly diagnosed a significant number of these 300 patients, and only two others had extra-chance success in the diagnosis of a single diagnostic group (in both cases, the brain-damaged group). The actual rate of success, however, ranged in correlational terms from .13 to .22 and was deemed far too small to be of use clinically.

It is concluded that although there is some nonchance relationship between Wechsler-Bellevue pattern and neuropsychiatric diagnosis, it is detectable for only some clinicians and then only to so small a degree as to be of little practical utility.

Received December 16, 1954.

References

1. Cohen, J. A factor-analytically based rationale for the Wechsler-Bellevue. *J. consult. Psychol.*, 1952, 16, 272–277.
2. Cohen, J. Factors underlying Wechsler-Bellevue performance of three neuropsychiatric groups. *J. abnorm. soc. Psychol.*, 1952, 47, 359–365.
3. Magaret, Ann. Intelligence testing and clinical practice. In L. A. Pennington & I. A. Berg (Eds.), *An introduction to clinical psychology*. New York: Ronald Press, 1948. Pp. 353–379.
4. Rapaport, D. *Diagnostic psychological testing*. Vol. I. Chicago: Year Book Publishers, 1945.
5. Wechsler, D. *The measurement of adult intelligence*. (3rd Ed.) Baltimore: Williams & Williams, 1944.
6. Wittenborn, J. R. An evaluation of the use of Bellevue-Wechsler subtest scores as an aid in psychiatric diagnosis. *J. consult. Psychol.*, 1949, 13, 433–439.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Specific Ability in IQ Change[1]

Charles T. Baker, Lester W. Sontag, and Virginia L. Nelson[2]
Fels Research Institute, Antioch College

Many investigators have found marked individual differences in the form of the mental growth curve when each individual's intelligence quotients or mental ages are plotted over a period of years.[3] Cycles of slow or rapid mental growth covering periods of long duration appear to be characteristic of large numbers of children from preschool years to adolescence. The consistency and magnitude of the trends of change in mental growth rate, repeatedly found in longitudinal studies (3, 4, 5, 6, 8, 10), indicate that these changes are not due to statistical artifacts. To date, the factors which contribute to accelerative and decelerative changes in mental growth rate have not been clearly defined.

In the study presented here we have investigated whether groups showing accelerated trends, decelerated trends, or little change in mental growth rate differ in their ability to pass the different types of Stanford-Binet items. Insofar as individual items of the Stanford-Binet require somewhat different abilities for their solutions, differences among groups in item difficulty provide information about group differences in areas of specific ability. The items of the Stanford-Binet have often been used in this diagnostic way. McNemar (9) suggested that groups of certain items could provide rough estimates of ability in nonverbal, memory, and vocabulary types of items.

Information regarding the group differences in ability to pass different types of items is also useful in determining the importance of a test-construction artifact in IQ changes over long periods of time. Since the Binet test includes more so-called nonverbal or performance type items at preschool levels and contains more verbal items at later ages, children with greater verbal ability could presumably score lower at earlier ages and higher at later ages as a function of the changing composition of the test itself, while the converse could be true of children whose best abilities were in the nonverbal area. This hypothesis has sometimes been given to account for the inconstancy of the IQ and the low predictive value of early tests (1, p. 294; 7, p. 643).

Procedure

The data for 140 children from the Fels Research Institute longitudinal study were used in the present analysis. Fifty-eight of the Fels subjects (*Ss*) admitted relatively early in the program received the 1916 Revised Stanford-Binet test at birthdays from age 3 until age 6; while 82 *Ss* admitted to the program in later years received Form L or Form M of the 1937 Revision at birthdays and six months after birthdays from age 2½ to age 6. From age 6 to 12 all 140 children received either Form L or Form M at birthdays. Form L and Form M were usually presented alternately. With few exceptions the tests were administered within a few days of scheduled birthdays or half-birthdays, and were administered by one examiner through-

[1] Presented in part at the Annual Meeting of the American Psychological Association, New York, September 1954.

[2] The authors gratefully acknowledge the many helpful suggestions of Dr. John I. Lacey and the statistical assistance of Mrs. Josephine D. Lubing.

[3] When one refers to measures of gain or loss in IQ trends over long periods of time this is equivalent to referring to measures of accelerative or decelerative changes in mental growth rate. Whether one chooses to graph IQ's or MA's is purely a matter of whether one prefers constancy to be represented by a straight line or by the diagonal line $y = x$. Accelerative and decelerative changes are depicted relative to the line which is chosen to represent constancy.

out the testing program. The mean IQ of the Fels children at any age is considerably above average on a variety of intelligence tests, the mean Stanford-Binet IQ for this sample being near 120 at all ages with standard deviations varying from 14 to 20 IQ points (10).

In order to divide the total group into experimental groups of those with accelerative changes, decelerative changes and constant mental growth rate trends, the following procedure was used. A smoothed longitudinal IQ curve was prepared for each *S* by the method of moving averages. Prior to age 6 smoothing was done by obtaining the mean of blocks of three consecutive raw scores when the interval between tests was six months, and by blocks of two when the interval between tests was one year. After age 6, blocks of three raw scores were averaged, the interval between tests being one year. The purpose of the smoothing was to eliminate erratic variations and thereby effectively portray general trends of IQ change. The different forms of the test were not taken into account in averaging the raw scores.

Quantification of the general trend in IQ change for each *S* was made in the following way. For the *Ss* who had been tested at yearly intervals prior to age 6 a "base IQ" was taken to be the mean IQ of raw score IQ's obtained at ages 3 and 4. For those who had been tested at six-month intervals the "base IQ" consisted of the mean IQ of tests given at ages 2½, 3, and 3½. For each *S* the amount of change in IQ from the "base IQ" to the highest smoothed IQ (in the case of those with accelerated trends) or to the lowest smoothed IQ (for those with decelerated trends) was determined. This yielded a gain or loss amount relative to the "base IQ." The *Ss* were then ranked from the individual gaining the most to the individual losing the most IQ points. The ranked sample of 140 *Ss* was then divided into experimental groups.

Group A consisted of the highest ranking 25 per cent of the sample, that is, those with greatest accelerative changes in mental growth rate. Group B, the 25 per cent of the sample who ranked in the middle of the distribution of ranks, was the group which showed least change; and group C was selected as the 25 per cent of the group who ranked as losing in IQ. It may be noted that only 75 per cent of the total sample was used. Approximately 13 per cent of the cases which ranked between group A and group B were eliminated from statistical treatment as were approximately 12 per cent of the cases ranking between groups B and C. These cases were discarded after graphic inspection of the individual IQ curves showed that changes in mental growth rate in these two parts of the ranked distribution were more erratic than many of the cases falling into the other groups.

The total sample was also subdivided in another dimension, being split into two matched groups (group I and II) of 70 each for purposes of cross-validation in this study and in future studies of these same *Ss*. Matching variables were date of birth, sex, level of IQ at age 3, and amount and direction of maximum IQ change. Matching was accomplished in such a way that groups A, B, and C were subdivided approximately equally so that statistical analysis of differences among groups A_I, B_I, and C_I could readily be cross-validated with groups A_{II}, B_{II}, and C_{II}.

A system devised by Baldwin (2) was used for estimating difficulty in first passing items of the Stanford-Binet, Form L. The difficulty level for an individual in first passing an item was defined as the mental age halfway between the mental age obtained by the individual at the time of his first success on the item and his mental age at the time of his previous failure on the same item. From this information about the difficulty of any item for each individual, it was possible to estimate the mean difficulty level of each item for the experimental groups.

Statistical analysis consisted of *F* tests of the differences in mean difficulty level among the experimental groups. This item analysis was done for 60 items of Form L of the Revised Stanford-Binet, and consisted of all the items from the test level of year IV through the year-XII series of items.

Results and Discussion

Seven of the 60 *F* tests of differences among groups A_I, B_I, and C_I were found to be significant below the .10 level in the first analysis. The seven items of the Stanford-Binet,

Form L, which resulted in these significant differences, were as follows: at year IV, items 2 and 6; at year IV–6, items 1, 3, and 6; at year V, item 2; and at year VI, item 5. None of the differences on these same items were significant upon replication with groups A_{II}, B_{II}, and C_{II}. In the second analysis 30 *F* tests were run on those items which had the greatest differences in mean difficulty level between groups A_I, B_I, and C_I. There were no trends toward significance between the first and second analysis. Since item difficulty was made relative to mental age, and therefore controlled for the effects of not only differential mental growth but also IQ level, it is concluded that there are no differences among the experimental groups in ability to pass different types of items. In view of the negative results on Form L, no attempt was made to analyze the data from Form M.

For the experimental purpose of this study it appeared to make no difference whether different forms or one form of the Stanford-Binet was used to obtain the smoothed trend line of IQ change. Smoothed IQ curves based on Form-L IQ's were prepared for those *Ss* who had received the 1937 Revision with Form L and Form M alternately presented. Using the procedure of ranking the individuals according to amount of change from the "base IQ" resulted in the *Ss* being placed in the same experimental groups as before. Smoothed curves based on Form-M IQ's only yielded similar results also. A comparison of the distributions of gains and losses was also made between those who had received the 1916 Stanford-Binet during the preschool years and those who had received the 1937 Revision throughout the testing period. There were no significant differences between these two groups in the mean amount of points gained or lost from the "base IQ" to the point of greatest divergence.

The maximum number of *Ss* possible in each experimental group in the first analysis was 18. The use of the Baldwin technique in estimating item difficulty level resulted in loss of data in the analysis of some items below the VII-year level of the test. Group A_I was combined with group A_{II}, B_I with B_{II}, and C_I with C_{II} for a nonreplicated design in testing two items when the number in one experimental group fell below 10. This analysis resulted in no significant differences in item difficulty.

Conclusions

Within the limitations of the sample and methods of this study it is concluded that there is no evidence of differences in ability to pass any type of Stanford-Binet item between those who accelerate in mental growth rate and those who decelerate. Accelerative or decelerative change in mental growth rate, at least in higher levels of IQ, is apparently associated with similar changes in growth rate in more specific areas of ability. There is no evidence that change in the rate of mental growth reflects a spurious gain or loss in IQ due to an artifact of test construction, wherein the ratio of verbal to nonverbal abilities changes with chronological age. These conclusions must, of course, be limited to the IQ level dealt with in this study.

It would appear that the factors which are associated with individual differences in the mental growth curve are more likely those factors which influence learning in any situation. Sontag and his associates (10) have presented data on the mental growth careers of the Fels subjects, and have proposed various aspects of motivation as determining factors in IQ change. Further research in this area promises to contribute knowledge of those factors which determine the form of the individual's mental growth curve.

Summary

The major purpose of this study was to determine whether *Ss* with accelerative changes in mental growth rate differed from those with decelerative changes in any specific areas of ability, as measured by the difficulty in first passing items of the Revised Stanford-Binet Scale, Form L. Matched experimental groups were obtained from 140 Fels *Ss* of normal to very superior intelligence, who had been tested at regular intervals from age 2½ to age 12. Cross-validation procedure was used in analyzing 60 items. There were no trends toward statistically significant differences between the groups in this analysis.

It is concluded that at least in the higher levels of intelligence, children with accelera-

tive changes in mental growth rate do not differ in any specific areas of intellectual ability from those with decelerative changes in mental growth rate. There was no evidence in this study to support a hypothesis that changes in Stanford-Binet IQ from the preschool years to later years are due to the changing ratio of verbal to nonverbal items in progressive age levels of the test.

Received December 16, 1954.

References

1. Anastasi, Anne, & Foley, J. P. *Differential psychology*. New York: Macmillan, 1949.
2. Baldwin, A. L. The relative difficulty of Stanford-Binet items and their relation to IQ. *J. Per.*, 1948, **16**, 417–430.
3. Bayley, Nancy. Mental growth in young children. *Yearb. nat. Soc. Stud. Educ.*, 1940, **39** (II), 11–47.
4. Bradway, Katherine. IQ constancy on the Revised Stanford-Binet from the preschool to the junior high school level. *J. genet. Psychol.*, 1944, **65**, 197–217.
5. Ebert, E., & Simmons, K. The Brush Foundation study of child growth and development. *Monogr. Soc. Res. Child Develpm.*, 1943, **8**, No. 2 (Whole No. 35).
6. Hilden, A. H. A longitudinal study of intellectual development. *J. Psychol.*, 1949, **28**, 187–214.
7. Jones, H. E. The environment and mental development. In Leonard Carmichael (Ed.), *Manual of child psychology*. New York: Wiley, 1954. Pp. 631–696.
8. Katz, E. The constancy of the Stanford-Binet IQ from three to five years. *J. Psychol.*, 1941, **12**, 159–181.
9. McNemar, Q. *The revision of the Stanford-Binet Scale*. Boston: Houghton Mifflin, 1942.
10. Sontag, L. W., Baker, C. T., & Nelson, Virginia L. Personality as a determinant of performance. Paper read at Amer. Orthopsychiat. Ass., New York, March, 1954.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Rigidity on the Rorschach and Response to Intermittent Photic Stimulation[1]

Laverne C. Johnson and John A. Stern

Department of Neuropsychiatry, Washington University School of Medicine and Research Laboratories, Malcolm Bliss Hospital

Most theories which attempt to account directly or indirectly for certain major phenomena of human behavior subscribe to a "unified" concept of personality. Such theories assume a high degree of constancy in the manner in which people perceive and respond to various situations. However, the experimental evidence on which such theories are based is meager and of recent origin (9, 11, 18). The experiment to be reported in this paper is, we hope, a contribution to the validation of one of the basic tenets of such theories of personality, namely, that a person will respond in a relatively constant manner to different perceptual situations. In addition, this study offers new opportunity to explore the mechanisms or processes which underlie responses to the Rorschach method as well as to show that it can be effectively used as an instrument for the study of perceptual processes (2, 10, 14).

The perceptual hypothesis to be investigated in this experiment is the following: A person who responds to one set of stimuli in a "rigid" or "flexible" manner will respond to other stimuli in a similar manner. While recognizing the importance of understanding the motivation of behavior, we are, at the present time, more concerned with determining that there are characteristic perceptual "styles" or "attitudes" which remain constant from one situation to another. Demonstration of the latter is a necessary first step toward the study of motivation.

In this study we are specifically concerned with the constancy of perceptual "flexibility" and "rigidity" from one ambiguous perceptual task to another. While well aware of the confusion surrounding the term "rigidity," we believe that it comes closer to describing one extreme of the process we have in mind than other less equivocal terms. By perceptual rigidity we mean that the person responds in a relatively restrictive or selective manner to environmental cues, as compared to the flexible person who responds in a less restrictive and selective manner to these same cues. Clinically, this type of person is often characterized as constricted and one who finds it necessary to deal with his environment in an unusually guarded and stereotyped manner.

The two perceptual tools utilized in this investigation are the Rorschach method and a photic stimulation experiment. In evaluating data from the Rorschach, Fisher's measure of "personality rigidity" was used (6). This measure is based on ". . . a variety of Rorschach signs which clinically have been found to characterize persons who are habitually restricted and who find it necessary to deal with the environment with an unusual amount of guarded caution" (6, p. 9). We feel that this definition, at least in part, describes what we mean by perceptual rigidity. Fisher's scale also has been found to be related to other measures of perceptual rigidity (1).

[1] This study was supported in part by a research grant from the National Institutes of Mental Health of the National Institutes of Health, United States Public Health Service.

These data were gathered under Air Force Contract Number 33(038)-13884 with the Department of Clinical Psychology, United States Air Force School of Aviation Medicine.

We wish to express our appreciation to Dr. George A. Ulett for making the data available and for his helpful suggestions in preparation of this paper.

The second ambiguous perceptual situation was intermittent photic stimulation at various frequencies. In this experiment the subject was requested to report what he saw after stimulation at various frequencies ranging from 3 to 30 flashes per second. These subjective sensations were the data used in this study.

In both tasks, the subject is asked to state what he experienced and what the stimulus material reminded him of or looked like. In both tasks the stimuli are ambiguous and permit a variety of responses.

In studies utilizing intermittent photic stimulation, Ulett and associates (15) reported various subjective sensations experienced by subjects. The content of these sensations ranged from pure description of colors and geometric designs to organized illusions and three-dimensional perceptions. Some subjects reported feelings of being projected into space, spinning, dizziness, anxiety, nausea, or other marked dysphoric experiences. Ulett found that it was possible to discriminate fairly well between nonanxiety-prone and anxiety-prone individuals according to the kind of sensation experienced during photic stimulation. More feelings of dysphoria were reported by anxiety-prone individuals.

As part of a larger study reported elsewhere (16), each subject was administered a Rorschach and his subjective sensations were recorded during intermittent photic stimulation. It was thus possible to obtain measures of perceptual rigidity for each task. Substantiation of the hypothesis that rigidity on the Rorschach is related to response to photic stimulation will add further support to the growing belief of constancy of perceptual "styles," as well as elicit new evidence that Rorschach responses reveal adaptive processes which are not different in kind from those appearing in other situations.

Subjects and Procedure

Subjects were 117 male students drawn from the entering class of a local Protestant seminary. The age range of the group was from 18 to 26 with a median of 20. For each subject (*a*) a battery of psychological tests, including the Rorschach, was administered, (*b*) an interview by two psychiatrists was given, and (*c*) an electroencephalogram (EEG) was recorded. Part of the EEG procedure included a photic stimulation survey of 3–30 flashes per second. There was a 40-second exposure to each of 21 different frequencies. Photic stimulation was produced by a pendulum-type episcotister interrupting a tungsten-filament light source which projected a 30-inch circle of light of 200 foot-candles intensity on an opal glass screen placed directly before the subject and filling his visual field. After each 40-second exposure, the subject was asked what he had seen or experienced. His responses were recorded as nearly verbatim as possible. These verbalizations were the data used in this study.

The psychological battery was the first procedure in the research; thus the Rorschach always preceded exposure to photic stimulation by at least a day. All the Rorschachs were administered and scored by the same person, an intern psychologist receiving training at Washington University School of Medicine after having completed all course requirements for the Ph.D. degree. The Klopfer scoring system (12) was used except for scoring of form level and originals where Hertz' (8) scores were used.

The measures of perceptual rigidity. As mentioned above, Fisher's rigidity scale was used as the measure of perceptual rigidity manifested in the Rorschach. The criteria for scoring the subjective sensations were a modification of those used by Klein and Schlesinger (10) in their study of perceptual attitudes toward instability. These criteria were developed and the subjective sensations were scored independently of the Rorschach measure of rigidity. The criteria used in this study were:

1. Criteria for perceptually rigid group.
   - *a.* Little verbalization; verbalization and elaboration does not increase as flicker frequency increases; verbalization may tend to decrease.
   - *b.* Responses are detailed descriptions of stimulus, e.g., geometric patterns or designs, lights, blocks, spots, circles, dots, squares, etc.
   - *c.* Movement seen predominantly descriptive and related to stimulus, e.g., whirling, shimmering, coming and going, up and down, fluctuates, fades in and out.

d. Uses same percepts, either color, form, or motion, and is not able to shift or elaborate on responses even when flicker frequency increases.

2. Criteria for the flexible group.
   a. Much verbalization and elaboration which tends to increase as flicker frequency increases.
   b. Responses not stimulus bound nor descriptive in nature, allows self to project forms and objects into stimulus.
   c. Movement seen also not stimulus bound nor predominantly descriptive, allows self freedom in expression and can take attitude of "distance" from stimulus.
   d. Content not stereotyped, uses more than one type of percept, allows self freedom to elaborate and change responses as flicker frequency increases.

Subjective sensations of each subject were judged as perceptually rigid or perceptually flexible by both authors and also were rated as to how well they met the criteria for that group. All ratings were made on a four-point scale of confidence extending from positive judgment that the subject falls into the given category to the feeling that the judgment was extremely tenuous. There was no fixed weighting procedure and it was rare that all the criteria for one group were met. The rating was in terms of the general over-all impression, and each judge was free to determine the importance of any one criterion. Disagreement between the two judges as to the group in which a subject should be placed eliminated the subject from the study. A difference of more than two points as to how well a subject met the criteria for the group in which he was placed also eliminated the subject. Fifteen subjects were eliminated because of difference as to group placement, and twelve more were eliminated because of difference of more than two points in the rating of certainty that the subject met the criteria. This reduced the *N* from 117 to 90. The over-all agreement between the two judges was 77 per cent, agreement for group placement alone was 87 per cent.

Fifty-six of our subjects were classified as rigid and thirty-four as flexible on the basis of their subjective sensation to photic stimulation.

Results

The results of the comparison of the rigidity score on the Rorschach and response to photic stimulation are presented in Table 1. These results clearly indicate that, in this sample, rigidity on the Rorschach is related to our measure of perceptual rigidity in response to photic stimulation. Those subjects who were placed in the perceptually rigid group in terms of their response to photic stimulation had significantly higher Rorschach rigidity scores. A biserial correlation was computed to obtain an estimate of the relationship between the two perceptual measures. This correlation was .59. If a Fisher Rorschach Rigidity score of 25 is used as a cutting point, eight of the flexible subjects (24 per cent) would be incorrectly placed in the perceptually rigid group. Fourteen of the perceptually rigid subjects (25 per cent) would be incorrectly labeled as perceptually flexible.

Table 1

Comparisons of Response to Photic Stimulation and Rorschach Rigidity Score

	Rorschach rigidity score				
Photic stimulation	*N*	Mean	*SD*	*CR*	*p*
Perceptually rigid	56	33.1	11.7	4.79	.0001
Perceptually flexible	34	20.8	12.8		

On the photic stimulation task a significantly larger number of subjects fell into the perceptually rigid than into the flexible group (56 rigid, 34 flexible). Two explanations seem probable. One, our criteria was at fault. Two, the larger number of perceptually rigid subjects was a function of the photic stimulation task. Both probably are factors. Our criteria were a first attempt at such discrimination, and refinement is undoubtedly possible. However, the most important factor seems to have been the general conditions under which our sample was subjected to the photic stimulation task.

For most of these subjects it was their first trip to a neuropsychiatric hospital where the laboratory is located, and it probably was disconcerting to have several needle electrodes stuck into one's scalp for the first time. There was no structuring as to what would happen to them or what they should see during photic stimulation. Thus it is not surprising that

many of the subjects were apprehensive, and reluctant and guarded about expressing themselves. Those able to express themselves freely are, perhaps, individuals who are able to adapt themselves to new and ambiguous situations.

The Rorschach, on the other hand, was administered in a more familiar situation—at the seminary attended by these students. The rigidity scores on the Rorschach were normally distributed, and the scores for both our perceptually rigid and flexible groups fall within one standard deviation of Fisher's normal group.

Comparison of Individual Rorschach Variables

To see which of the determinants were contributing most heavily to the differentiation between the two groups, a comparison of individual Rorschach variables was made. Cronbach (3) has pointed out many of the errors and pitfalls in such comparisons. Of particular importance is the positive relationship between the number of responses (R) and a number of clinically important determinants. Because the number of responses were significantly different between our two groups (the perceptually flexible group gave more responses than the perceptually rigid group), the chi squares for number of movement ($\#M$) and number of original responses ($\#O$) are reported with and without corrections for R. The method of correction was that suggested by Cronbach (3, p. 411). As most of the distributions were skewed, chi-square measures of significance were computed. The median of the combined groups was used as the dividing point. A contingency coefficient was computed to give an estimate of the degree of association between the two groups for each variable. These correlations, of course, are not comparable to product-moment correlations and are limited by the number of categories, but their use does not include assumptions of normality or linearity.

Of the 17 Rorschach variables for which chi squares were computed, five were significant beyond the .01 level as indicated in Table 2 ($\#R$, $\#M$ corrected and uncorrected for R, $\#O$ uncorrected for R, T/R, $T/1R$). Those not significant were $W\%$, $D\%$, $Dr\%$, $S\%$, $F\%$, $\#FC$, $\#CF + C$, *Sum C*, $\#O$ corrected for R, $\#P$, $P\%$ and variety of content. As O when corrected for R was not significant, the difference in number of original and variety of content responses seems to be a function of difference in R.

These findings relate well to measures of personality rigidity used in the clinical evaluation of the Rorschach. Most Rorschach experts would agree that restriction in number of responses, few movement responses, greater time per response and delayed reaction time are major signs utilized in the determination of personality rigidity. The number of original responses is also often used as an indicator of rigidity, and even though our flexible groups gave significantly more originals, the difference failed to be significant when corrected for R.

The failure of $F\%$ and $F + \%$ to be significantly different in the two groups was surprising. $F + \%$ approached significance, p of .09, and was in the expected direction, but in light of the common use of $F\%$ and $F + \%$ as indicators of constriction and rigidity, we expected the difference to be larger.

Because of the differing reports (3, 5, 13) on the relationship of M to R, a correlational scattergram was plotted and as the relation-

Table 2

Comparison of Individual Rorschach Variables

Rorschach variable	Mean scores perceptually Flexible	Rigid	Chi square	p	C
R	48.3	34.5	11.53	.001	.34
$\#M$ (uncorrected for R)	4.62	2.28	20.79	.001	.44
$\#M$ (corrected for R)			7.69	.008	.28
$\#O$ (uncorrected for R)	10.2	6.1	10.76	.001	.33
$\#O$ (corrected for R)			.027	NS	—
T/R	36.1	53.9	17.80	.001	.41
$T/1R$	18.5	27.6	8.30	.006	.29

Table 3

Comparison of Groups Using a Plus or Minus Scoring Procedure and Comparison of Individual Elements in the Fisher Scale

Individual elements	Per cent plus scores perceptually Flexible	Rigid	*CR*	*p*
All elements	33.9	43.2	2.67	.008
F%	41.1	46.4	.49	—
#R	3.0	19.6	2.26	.02*
A%	8.8	5.4	.51	—*
F+%	35.3	55.4	1.66	.06
Color	52.9	57.1	.39	—
W	23.5	33.9	1.05	—
Dd	38.2	42.9	.48	—
M	17.6	51.8	3.23	.001
Content	76.5	64.3	1.22	—
T/1R	23.5	35.7	1.22	—
Shading responses	52.9	62.5	.90	
#R+F+%+M	18.6	42.3	4.16	.0001

* Corrected for continuity.

ship between *M* and *R* was linear, a Pearson product-moment correlation was computed. The correlation was .50. The correlation between *#O* and *#R*, similarly computed, was .83. It is thus not surprising that our chi-square value for *M* should decrease from 20.79 to 7.69, and from 10.76 to .027 for *#O* when corrected for *R*. These findings lend support to the findings of Fiske and Baughman (5) and indicate the importance of controlling *R* when using individual Rorschach determinants for comparative purposes.

Analysis of Elements in Fisher's Rorschach Scale

The analysis of the individual Rorschach variables and inspection of the data suggested that all elements in Fisher's Rorschach scale are not contributing equally to the total score, and that a differentiating measure of Rorschach rigidity was possible by utilizing a simple plus or minus scoring procedure. In the usual procedure a weighted score is assigned as a penalty for what is considered to be excess rigidity or restrictiveness. The larger the final summation of weights, the greater is the implied rigidity. In this present analysis, instead of assigning a weighted score, only a plus was given when a penalty score was indicated. When no penalty score was indicated, a minus was given. The individual elements of the Fisher scale were similarly analyzed with the exception of card turning which was not scored because of incompleteness of data. The percentage of plus signs for each group, and the differences are given in Table 3.

As expected, the over-all difference between the two groups is highly significant. Those classified as perceptually rigid on the basis of their response to photic stimulation have a significantly larger percentage of plus scores than do those labeled perceptually flexible. The results of the analysis of individual elements are consistent with those found in the comparison of individual Rorschach variables. Human movement and number of responses were significant beyond the .05 level, and *F + %* at the .06 level. While the other scores were not significantly different, the differences were in the expected direction with the exception of *A%* and content. The use of only *#R*, *F + %* and *M* in future research of this kind is suggested by the highly significant difference obtained when only these three elements are used for comparison.

Relation of Perceptual Rigidity—Flexibility to Intelligence and Anxiety Proneness

The results on the Rorschach suggest that the perceptually flexible group is possibly more intelligent than the perceptually rigid one, raising the possibility that the life styles in question are simply those of more or less intelligent students. To evaluate this possibility two measures of intelligence were utilized for comparison, ACE scores and a short form of the Wechsler-Bellevue (Comprehension, Vocabulary and Similarities). The two groups did not differ significantly on either the group or individual test. The mean ACE score for the rigid group was 132.8 and 135.8 for the flexible group. The mean CVS scores were 41.7 and 42.4 for the rigid and flexible groups, respectively. Thus the possibility that the style of behavior exhibited in these two perceptual situations is a function of intelligence is not supported.

As the psychological tests and psychiatric interviews of which these data are a part were used extensively to judge anxiety proneness, a comparison between the anxiety-proneness score and our measures of rigidity-flexibility was possible. This seemed to be particularly

relevant since Ulett (16) has found this measure of anxiety proneness to be related to feelings of dysphoria during photic stimulation and since there are some who think of perceptual rigidity as a result of anxiety or a defense against anxiety (4).

Neither of the measures of perceptual rigidity-flexibility was significantly correlated with anxiety proneness as measured in the original study. The Pearson product-moment correlation between Fisher Rorschach rigidity score and anxiety proneness was .10, and the biserial correlation between ratings of perceptual rigidity-flexibility based on the subjective response to photic stimulation and anxiety proneness was .11.

This low relationship raises new questions as to the relationship of rigidity-flexibility to anxiety. Many readers may draw the conclusion that the anxiety measure was inadequate. This cannot now be ruled out. However, the rating of anxiety that was used was based on a number of the commonly accepted psychological measures of anxiety in use today. The Rorschach, Taylor Anxiety Scale, Saslow Screening Questionnaire, Digit Span, Mirror Drawing, Draw-A-Person, evaluation of general test behavior, and psychiatrists' ratings all were utilized in evaluating anxiety proneness.

Three other alternatives seem possible. First, the relationship between anxiety and perceptual rigidity is complex, that is, perceptual rigidity may be a successful defense against anxiety, and, therefore, some highly rigid individuals would show little or no anxiety. It is interesting that Fisher found hysterics and paranoids, two clinical groups who show the least manifest anxiety, to be the most rigid. Certainly this whole area of perceptual rigidity, supposedly the result of anxiety, is poorly defined, unclear, and needs further study.

A second alternative is that perceptual rigidity-flexibility is an "autonomous" or "conflict-free" function of the ego as discussed by Hartman (7). That is, it is a function not exclusively or necessarily developed from or in the service of drive conflicts. Klein and his co-workers (11) believe that their results seem increasingly relevant to the above formulation. They state: "It seems a worthwhile hypothesis that those aspects of perceptual process which psychoanalytic theorists currently view as 'autonomous' are themselves idiosyncratically organized within people, i.e., into 'styles of autonomous functioning' " (11, p. 3).

A third alternative suggested by Gray Walter's work (17) is that the behavior exhibited is a resultant of physiological factors not yet determined rather than being a function of anxiety. He has reported that the onset of a particularly vivid illusion or hallucination during photic stimulation was preceded or accompanied by a characteristic alteration in the electrical response of the brain. Emotional and autonomic disturbances were usually correlated with augmented response at 6–7 cycles per second from the parieto-temporal region. The use of electronically analyzed EEG records during photic stimulation may help in indicating the association of subjective sensations to electrical discharge in the brain even though no cause-and-effect relationship can be inferred at this time. Further work is in progress in this area.

Summary and Conclusion

The results of this study add new evidence to the growing belief that there are "perceptual styles" or "perceptual attitudes" which are consistent modes of response, and that the Rorschach method can be used effectively as a perceptual instrument in the study of personality. We have shown that individuals who respond in a perceptually rigid manner in one perceptual task will respond in a like manner to a different perceptual situation. The one factor which both tasks have in common is that they are ambiguous perceptual situations. The hypothesis was proposed that individuals respond in a relatively consistent manner to environmental cues, and two perceptual tasks were utilized to test this. We further hypothesized that the Rorschach method could be used as a tool for the study of perception and that Rorschach responses would reveal adaptive processes not different in kind from those appearing in other perceptually ambiguous situations. To test the hypotheses we utilized data collected on 117 freshman students from a local seminary. These students were given the Rorschach, and their sub-

jective sensations during intermittent photic stimulation were recorded. The subjective sensations were used for comparison with Rorschach responses.

The Rorschachs were scored for rigidity according to the technique developed by Fisher. The subjective sensations were scored independently utilizing an a priori criterion for dividing the sample into two groups, perceptually rigid and perceptually flexible.

The perceptually flexible group had significantly lower Rorschach rigidity scores than did the perceptually rigid group. When the individual Rorschach determinants were examined, the perceptually flexible group was found to have given significantly more responses, more movement responses, taken less time per response, and had a significantly lower average reaction time. In contrast to findings reported to date, $F\%$ and $F + \%$ did not significantly differentiate the two groups. Neither of the two perceptual rigidity-flexibility measures was related to intelligence-test scores or to a measure of anxiety proneness. The significance of these findings is discussed and alternative hypotheses are suggested.

Received January 6, 1955.

References

1. Becker, W. C. Perceptual rigidity as measured by aniseikonic lenses. *J. abnorm. soc. Psychol.*, 1954, **49**, 419–426.
2. Bruner, J. S. Perceptual theory and the Rorschach Test. *J. Pers.*, 1948, **17**, 157–168.
3. Cronbach, L. J. Statistical methods applied to Rorschach scores. *Psychol. Bull.*, 1949, **46**, 393–429.
4. Ericksen, C. W., & Eisenstein, D. Perceptual rigidity and the Rorschach. *J. Pers.*, 1953, **21**, 386–391.
5. Fiske, D. W., & Baughman, E. E. Relationship between Rorschach scoring categories and the total number of responses. *J. abnorm. soc. Psychol.*, 1953, **48**, 25–32.
6. Fisher, S. Patterns of personality rigidity and some of their determinants. *Psychol. Monogr.*, 1950, **64**, No. 1 (Whole No. 307).
7. Hartman, H. The mutual influences in the development of ego and id. In *Psychoanalytic study of the child*. Vol. VII. Ruth Eissler, Anna Freud, H. Hartman, & E. Kris (Eds.). New York: International Universities Press, 1952.
8. Hertz, Marguerite W. *Frequency tables for scoring responses to the Rorschach Ink Blot test*. Cleveland, Ohio: Western Reserve Univer. Press, 1952.
9. Klein, G. S. The Menninger Foundation research on perception and personality. *Bull. Menninger Clin.*, 1953, **17**, 93–99.
10. Klein, G. S., & Schlesinger, H. J. Perceptual attitudes toward instability: I. Prediction of apparent movement from Rorschach responses. *J. Pers.*, 1951, **19**, 289–302.
11. Klein, G. S., Murphy, G., Gardner, R. W., Holzman, P. S., & Schlesinger, H. J. The relation of consistencies in cognitive behavior to personality organization. 1953–54 Progress Rep. on Proj. MH-25(C). On file, National Institutes of Health, Bethesda, Md.
12. Klopfer, B., & Kelley, D. G. *The Rorschach technique*. New York: World Book, 1946.
13. Kornetsky, G., & Gerard, D. L. Effect of increasing the number of Rorschach responses on *Sum C* and *M*: A note on Fiske and Baughman's study. *J. abnorm. soc. Psychol.*, 1954, **49**, 592–593.
14. Matarazzo, Ruth, Watson, R. I., & Ulett, G. A. Relationship of Rorschach scoring categories to modes of perception induced by intermittent photic stimulation. A methodological study of perception. *J. clin. Psychol.*, 1952, **8**, 368–374.
15. Ulett, G. A., Gleser, Goldine, Winokur, G., & Lawler, Ann. The EEG and reaction to photic stimulation as an index of anxiety proneness. *EEG clin. Neurophysiol.*, 1953, **5**, 23–32.
16. Ulett, G. A., Gleser, Goldine, Starr, P., Haddock, J., Lingley, L., & Lawler, Ann. Psychiatric screening of flying personnel. Further studies toward the development of an electroencephalographic screening technique. Proj. Rep. Number 5, School of Aviat. Med., San Antonio, Texas, 1953.
17. Walter, U., & Walter, W. G. The central effect of rhythmic sensory stimulation. *EEG clin. Neurophysiol.*, 1949, **1**, 57–86.
18. Witkin, H. A., Lewis, H. B., Hertzman, M., Machover, Karen, Meissner, Pearl, & Wapner, S. *Personality through perception*. New York: Harper, 1954.

NEW BOOKS AND TESTS

Books

Allport, G. W. *Becoming: basic considerations for a psychology of personality*. New Haven, Conn.: Yale Univer. Press, 1955. Pp. ix + 106. $2.75.

Revised from Yale's Terry Lectures, which challenge the lecturer to relate his specialty to human welfare and to a broad conception of religion, this little book is a quietly bold attempt to present a perspective in psychology on such matters as freedom, the religious sentiments, and the democratic ideal. The unifying concept is that of personality, taken as something continuously *becoming* rather than ever meaningfully identifiable on cross-sectional analysis. Personal uniqueness, selfhood, and autonomy, familiar Allportian principles, play their predictable roles in the emphasis on the person as the growing and evolving unit of psychological interest. Finally, psychology's prime task is conceived as that of determining whether "the democratic ideal is viable." Read as an impressively learned and even poetically sensitive statement of a possible relationship between psychological science and certain vital human values, *Becoming* provides a profiable experience for any psychologist concerned about either his discipline or his status as an intellectual in the republic's citizenry. Similarly, Allport's insistence on the process character of personality is useful and welcome. One may wonder, however, if such borrowings as Swedenborg's concept of the proprium is not evidence for a flirting with reification in the ideas of personality and the self, and if the avowed acceptance of Leibnitzian idealism has not led to a kind of stubborn misunderstanding of both psychoanalysis and neobehaviorism. Both seem to serve only as convenient but ghostly whipping boys. Even among those who share Allport's values and admire his humanity, there is still room to doubt that his methodological point of view does other than inhibit the kinds of understandings for which he argues so urgently and with such cultivated persuasion.—*E. J. S.*

Arieti, Silvano. *Interpretation of schizophrenia*. New York: Robert Brunner, 1955. Pp. xxiii + 522. $6.75.

In spite of the author's modest disclaimers, *Interpretation of Schizophrenia* is a thorough, systematic treatment, and about as complete as a single volume on so vast a subject can be expected to be. Dr. Arieti proceeds logically from a historical development to psychodynamics, through a study of formal mechanisms in schizophrenia, to discussions of progressive stages of the disease, the organic manifestations, and methods of treatment. The point of view is dynamic without an underestimation of somatic factors. Of special interest to psychologists are chapters on "The Retreat from Emotions," "The Retreat of Reason," and "The Retreat from Sanity," to single out only three of the twenty-four excellent chapters. The book is well-written and reads easily; case discussions are few but rather full, although illustrative behavior samples abound; a bibliography of 280 titles is listed. It is an important reference work as well as a balanced text on schizophrenia.—*M. K.*

Blum, Lucille H., Davidson, Helen H., & Fieldsteel, Nina D. *A Rorschach workbook*. New York: International Universities Press, 1954. Pp. 166 (paper). $2.00.

This lithographed paper-bound volume introduces the student to each of Klopfer's major scoring determinants and provides an extensive series of scoring exercises for each. The "correct" answers are provided in an Appendix. A brief final chapter is devoted to administration with special attention to the inquiry as an essential prerequisite to accurate scoring. Neither the text nor the exercises are concerned with interpretation, which in the authors' opinion should follow the development of skill in scoring. The "basic bibliography on scoring" has seven titles.—*E. L. K.*

Patrick, Catharine. *What is creative thinking?* New York: Philosophical Library, 1955. Pp. xi + 210. $3.00.

This is an amalgam of well-known publications on creative thinking, with selected materials on reasoning. Included also are a chapter on age and productivity and a program for improving creative thinking. The quotations and data are organized around stages of preparation, incubation, illumination, and revision or verification. There is a comprehensive bibliography.—*A. R.*

Rose, Arnold M. (Ed.) *Mental health and mental disorder, a sociological approach*. New York: Norton, 1955. Pp. xiv + 626. $6.50.

This volume is the first of a projected series to be published under the auspices of the Society for the

Note.—The reviews were prepared by the Editor and the Advisory Editors, who may be identified by their initials.

Study of Social Problems, an interdisciplinary group concerned with the integration of sociological, psychological, and anthropological research and the application of these researches to human problems. It contains a collection of thirty-eight essays, case studies, research reports, and systematic analyses of the literature, planned to give an overview of the rather scattered materials in the field. Ten of the papers were prepared especially for the volume; the others have appeared previously, chiefly in sociological journals. The papers are presented in seven groups, the titles of which indicate the wide scope: problems of social psychiatry and theoretical overview, social characteristics of the mentally disordered, mental disorder in the community, sociological approaches to the study of specific disorders, social psychology of personality organization and disorganization, the sociological approach to problems marginal to mental disorder, and contributions to the understanding of mental health. The survey may be useful to psychologists. It concentrates a considerable amount of data and discussion, and most of the papers have extensive bibliographies.—*A. R.*

Royce, James E. *Personality and mental health.* Milwaukee, Wis.: Bruce, 1955. Pp. xiv + 352. $3.50.

This brief survey of modern clinical psychology is scholastic in viewpoint and "integrated throughout with a frankly Christian philosophy of life." The author planned his basic text in mental hygiene and the psychology of adjustment with reference to the problems of students. He therefore did not attempt to include references on all the material or on the various areas, but gave emphasis to classroom situations and students' needs.—*B. M. L.*

Ulett, George. *Rorschach introductory manual.* Beverly Hills, Calif.: Western Psychological Services, 1955. Pp. 45 (paper). $3.50.

This is the second edition of a manual first published in 1949 as an aid in teaching the Rorschach technique to physicians. It is an attractively printed lithographed booklet, covering the essentials of administration, scoring, and interpretation. Included also are a percentage calculation table, an ingenious multicolored "Interpretation Chart," and a table of "signs" to aid in clinical diagnosis. The approach is eclectic, and the text, while didactic, is extremely lucid. The first sentence in the text states: "The Rorschach test is a remarkably accurate means of gaining information about a patient's personality . . ."; later, in the chapter on interpretation, the author warns that "some of the Rorschach theory and interpretation given in the following pages are still open to question." There is a selected bibliography of 61 titles.—*E. L. K.*

Tests

Holzinger, Karl J., & Crowder, Norman A. *Holzinger-Crowder Uni-Factor Tests.* Grades 7–12. 2 forms. 40½ (90) min. IBM or hand scoring. Test booklets ($5.70 per 35) with manual, pp. 30; 2 answer sheets ($1.80 per 35, and $1.15 per 35); keys (50¢); specimen set (50¢). Yonkers, N. Y.: World Book Co., 1955.

The *Uni-Factor Tests* consist of nine subtests which measure four factors: verbal, spatial, numerical, and reasoning. Each factor is scored separately, and interpreted by the aid of percentile tables which give norms for each grade from 7 to 12. In addition, a Scholastic Aptitude Score may be obtained from a simple regression equation which combines the verbal, numerical, and reasoning factor scores. Alternate-form and split-half reliabilities for the separate factors compare favorably with other similar instruments. Typical intercorrelations among the factors, for single-grade groups, range from about .30 to about .50, and show that the scores, while not independent, possess considerable distinctiveness. The Manual contains unusually full data on the relationships of the factors to other concurrent measures—other intelligence tests, achievement tests, and school marks. In these data, differential values of the factors are evident. Although the ultimate merit of the test in predicting academic and vocational criteria remains to be determined, it represents one more thoughtfully constructed tool for analyzing "global intelligence" into some meaningful parts.—*L. F. S.*

Wechsler, David. *Wechsler Adult Intelligence Scale (WAIS).* Ages 16 and above. Individual test. 1 form. Test material, with 25 record forms, and manual, pp. vi + 110 ($21.00); record form only ($1.60 per 25); manual only ($2.75). New York: Psychological Corp., 1955.

The long-awaited revision of the Wechsler scale will fulfill the demands of psychologists—it is enough of a revision to remedy the major shortcomings of the 1939 Wechsler-Bellevue, and not so sweeping a change as to discard the experience gained by clinicians from the use of the older test. The eleven subtests remain the same, but all except the Digit Span have been revised in some degree. The Vocabulary subtest is wholly new; troublesome and outdated items have been removed from Information, Comprehension, and Arithmetic, and the Object Assembly has been lengthened by adding another object, an elephant. The major standardization from ages 16 to 64 was based on 1,700 examinees with equal numbers of men and women, and with a precise stratification of geographical region, urban-rural residence, race, and occupation based on the 1950 census. The manual contains appropriate data on reliabilities and subtest intercorrelations for three age segments. Corrected split-half reliabilities are .97 for the full scale and range from .66 to .96 for the subtests. The manual and record blank both warn against making clinical interpretations from profiles

or from differences between subtest scores with limited reliability. Left for future research, perhaps quite properly, are the relationships between the old and new scales, and the validities of the revised test for various applications. The test itself and the data in the manual give the impression that the WAIS is a mature and sophisticated instrument, a worthy successor to the Wechsler-Bellevue for both clinical service and research.—*L. F. S.*

Wittenborn, J. Richard. *Wittenborn Psychiatric Rating Scales.* Psychiatric patients. Individual rating scale. 1 form. Rating scales ($2.75 per 25); with manual, pp. 12; specimen set (35¢). New York: Psychological Corp., 1955.

A series of 18 research articles since 1950 has reported the development and application of the Wittenborn scales. Now they are made available for general use in an attractive and convenient format. The scales are an aid to the objective description of psychiatric patients in terms of their observable symptoms. On each of 52 areas of behavior, a patient is rated by selecting the one of four statements which describes him best. The quantified ratings are turned into scores on nine factors identified by factor analytic studies. The corrected split-half reliabilities of the factor scores range from .67 to .92, with only one scale below .76; their intercorrelations are generally low. The new blank greatly simplifies the calculation of the factor scores by reducing it to the addition of digits in conveniently arranged columns and the use of a simple table. With their new convenience and availability, the scales will probably receive wider use in research. They make a valuable contribution to the knottiest problem in studies of psychopathology, the problem of the criterion.—*L. F. S.*

Books Received

Bettelheim, Bruno. *Truants from life.* Glencoe, Ill.: Free Press, 1955. Pp. xvi + 511. $6.00.

Bram, Joseph. *Language and society.* Doubleday Short Studies in Sociology, No. 8. Garden City, N. Y.: Doubleday, 1955. Pp. viii + 66 (paper). 95¢.

Chinoy, Ely. *Automobile workers and the American dream.* Garden City, N. Y.: Doubleday, 1955. Pp. xx + 139. $3.00.

Deutsch, Felix, & Murphy, William F. *The clinical interview.* Vol. II: *Therapy.* New York: International Universities Press, 1955. Pp. 335. $7.50.

Faris, Robert E. L. *Social disorganization.* (2nd Ed.) New York: Ronald, 1955. Pp. viii + 664. $5.50.

Frank, Lawrence K. *Individual development.* Doubleday Papers in Psychology, No. 14. Garden City, N. Y.: Doubleday, 1955. Pp. vii + 52 (paper). 85¢.

Freeman, Frank S. *Theory and practice of psychological testing.* (Rev. Ed.) New York: Holt, 1955. Pp. xvi + 609. $5.25.

Ghiselli, Edwin E., & Brown, Clarence W. *Personnel and industrial psychology.* (2nd Ed.) New York: McGraw-Hill, 1955. Pp. ix + 492. $5.00.

Gillespie, James M., & Allport, Gordon W. *Youth's outlook on the future.* Doubleday Papers in Psychology, No. 15. Garden City, N. Y.: Doubleday, 1955. Pp. ix + 61 (paper). 85¢.

Glover, Edward. *The technique of psycho-analysis.* New York: International Universities Press, 1955. Pp. x + 404. $7.50.

Greer, Scott A. *Social organization.* Doubleday Short Studies in Sociology, No. 9. Garden City, N. Y.: Doubleday, 1955. Pp. x + 68 (paper). 95¢.

Jewish Vocational Service. *Words for work.* Boston: Jewish Vocational Service, 1955. Pp. 140 (paper). $1.00.

Karn, Harry W., & Weitz, Joseph. *An introduction to psychology.* New York: Wiley, 1955. Pp. xii + 315. $3.90.

McCary, James L. (Ed.) *Six approaches to psychotherapy.* New York: Dryden Press, 1955. Pp. vii + 402. $3.75.

Maier, Norman R. F. *Psychology in industry.* (2nd Ed.) Boston: Houghton Mifflin, 1955. Pp. xiv + 678. $5.50.

Mayer, Jurt B. *Class and society.* Doubleday Short Studies in Sociology, No. 10. Garden City, N. Y.: Doubleday, 1955. Pp. vii + 88 (paper). 95¢.

Michaels, Joseph J. *Disorders of character.* Springfield, Ill.: Charles C Thomas, 1955. Pp. x + 148.

Munn, Norman L. *The evolution and growth of human behavior.* Boston: Houghton Mifflin, 1955. Pp. xi + 525. $6.50.

National Conference of Social Work. *Casework papers, 1954.* New York: Family Service Association of America, 1955. Pp. 150 (paper).

Oates, Wayne E. *Religious factors in mental illness.* New York: Association Press, 1955. Pp. xx + 239. $3.50.

Podolsky, Edward. (Ed.) *Management of addictions.* New York: Philosophical Library, 1955. Pp. xvii + 413. $7.50.

Ralli, Elaine P. (Ed.) *Adrenal cortex.* Transactions of the Fifth Conference, 1953. New York: Josiah Macy, Jr. Foundation, 1954. Pp. 187. $3.75.

Reitman, Francis. *Insanity, art, and culture.* New York: Philosophical Library, 1954. Pp. 111. $3.75.

Rothstein, Jerome H., & O'Connor, Thomas. *Films on the handicapped.* Washington: International Council for Exceptional Children, 1955. Pp. vi + 56 (paper). $1.00.

Thorpe, Louis P. *Child psychology and development.* New York: Ronald, 1955. Pp. v + 709. $6.00.

U. S. Department of Health, Education, and Welfare. Public Health Service. *Evaluation in mental health.* Washington: U. S. Government Printing Office, 1955. Pp. x + 292. $2.00.

Wallin, J. E. Wallace. *Education of mentally handicapped children.* New York: Harper, 1955. Pp. xiii + 485. $4.50.

Wallin, J. E. Wallace. *The odyssey of a psychologist.* Wilmington 4, Del. (311 Highland Ave., Lyndalia): Author, 1955. Pp. xvii + 243 (paper). $3.00.

Again Available

SYMPOSIUM: PSYCHOANALYSIS AS SEEN BY ANALYZED PSYCHOLOGISTS

This well-known symposium, which first appeared in the *Journal of Abnormal and Social Psychology* in 1940, has now been reprinted. The contributors are Edwin G. Boring, Carney Landis, J. F. Brown, Raymond R. Willoughby, Percival M. Symonds, Henry A. Murray, Else Frenkel-Brunswik, David Shakow, Donald V. McGranahan, and Austin B. Wood. Comments on the symposium are made by Hanns Sachs and Franz Alexander. 160 pages.

Price, $2.00

Order from

American Psychological Association
1333 Sixteenth Street N. W.
Washington 6, D. C.